Promoting Representation Through Data: The Case for More Comprehensive Ethnicity Data in Australia

Liz Allen Demographer and lecturer at the ANU Centre for Social Research and Methods

Australian National University, Canberra ACT, Australia

ABSTRACT

What if data used to inform knowledge is incomplete or false, leading to misunderstandings about the social world? In Australia, what is known about ethnic diversity is based on an outdated definition and inadequate measure of multiculturalism. This article explores how gaps in data can lead to and further entrench disempowerment. Using a sociodemographic approach, this paper examines the ways in which data can create and maintain poor representation; via collection, analysis and infrastructure. This paper demonstrates that what is counted matters for equality, and lays out what is necessary to help promote ethnic diversity through data collection.

Keywords - data, data infrastructure, ethnicity, census, inequality

---

Disclosure statement: No potential conflict of interest was reported by the authors.

License: This work is under Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Suggested citation: Allen, L. (2021). "Promoting representation through data: the case for more comprehensive ethnicity data in Australia." *Law in Context*, 37 (2): 54-61, DOI: <https://doi.org/10.26826/law-in-context.v37i2.144>

Summary

1. Introduction
2. Background and context
  - 2.1. *Legacy of a white Australia on data*
  - 2.2. *Data landscape*
3. Better capturing ethnic identity to empower
4. Conclusion and discussion
5. References

1. INTRODUCTION

Data is the foundation upon which knowledge is built. But what if the data which informs what is known is incomplete or false? The meanings individuals derive from the social world come from the interpretations of

a multitude of lifetime observations and experiences, and frame the way in which people interact and position themselves in society. Humans tend to absorb meaning through interactions and from observations like data

points held in an internal file within the brain, and it becomes a hard-wired visceral process of sorts, shaping individual perceptions which become more powerful than reality. Indeed, much of what we know, or think we know, about the world around us comes from our individual observations and internal recordkeeping. Everything from the level of neighbourhood safety to fears of migrants stealing local jobs, perceptions shape the way people not only view the world but, more importantly, interact within it. For example, parents who report higher levels of perceived neighbourhood crime and disorder tend to limit their children's activity within the built environment, with real and sometimes lasting health impacts from reduced physical activity among young people (Timperio et al. 2005).

Poor attitudes towards migrants, which have been historically framed by protectionist government policies, tend to be based on intangible fears held by the dominant group concerning loss of social standing rather than the plethora of evidence to the contrary (Allen 2020). These examples highlight how pernicious individual perception formation can be in creating a sense of evidence from anecdotes (a data of sorts): I see it, therefore it is.

The endeavor to provide data and commentary to dispel myths and thereby question a traditional or colloquial way of knowing seems fruitless given the challenges posed by lived experience. A post-truth environment undoubtedly makes the efforts to apply data and evidence to guide policy and practice ever more difficult (Nash 2020). The elevation of so-called "alternative facts" as a peculiar corollary to recognized empirical evidence emboldens the denial and erosion of data and evidence (Morrissey 2017).

There is much potential success in the use of data and evidence to promote fairness and equality. The same desire for countering facts and evidence with alternative narratives of reality can be used to promote the collection of data to fill gaps in knowledge. Indeed, there is a thirst for data. James Baldwin, when discussing the civil rights movement in the United States in the 1960s, famously said: "not everything that is faced can be changed, but nothing can be changed until it is faced"

(Yu 2017). Without data, conversations concerning what is, and could be, lack power — data elevates and exposes topics that have been ignored, silenced or overlooked due to a lack of evidence or knowledge of their very existence. Data can, and ought to, be used to keep systems fair and reflect society (Federation of Ethnic Communities' Council of Australia 2020).

This paper considers the question: what if evidence informing the knowledge about multiculturalism in Australia is wrong? This question guides the exploration here of the limitations of what can be known about ethnic identity in Australia versus what we think is known. Further, this data incongruence is proposed to increase the risk of inequality through a lack of appropriate representation in data.

Importantly, this paper considers ethnic diversity generally and does not explicitly explore the recognized intersection of ethnicity and that of First Nations identity in Australia.

2. BACKGROUND AND CONTEXT

Australia is considered among the most multicultural nations in the world, with nearly one-third of Australians born overseas (ABS NDa). This statistic is based on a simple reduction of background of individuals, based on the outdated notion of race—country of birth—rather than the now more normative social concept of ethnicity. Ethnicity "is defined as a sense of belonging, based on ones' ancestry, cultural heritage, values, traditions, rituals, and often language and religion" (Green et al. 2015, p. 676).

In countries open to immigration, like Australia, birthplace tells only part of a more complicated story of the ways in which people identify, articulate and perform their ethnicity (Rocha et al. 2018). The focus on birthplace hides a much broader scale of multiculturalism that exists in Australian society (Khoo 1991). A higher degree of multiculturalism, hidden from data, has the potential to disempower those who identify as something other than the majority.

Religion, ancestry, parental country of birth, and language are other measures used in Australia to

quantify ethnic diversity. However, these indicators are considered surrogate measures of ethnicity and inadequate for capturing the true nature of diversity in Australia, especially as mixed partnering (interracial relationship) increases (Horn 1987). There have been growing attempts to solve the issue of hidden diversity in Australian statistics, but the history of British colonization and the data landscape appear to get in the way of progress on this front (Federation of Ethnic Communities' Council of Australia 2020).

2.1 LEGACY OF A WHITE AUSTRALIA ON DATA

Notions of ethnicity in Australia are framed by the history of colonization and the subsequent (yet enduring sentiments of) White Australia policy (Allen 2020). As a result, the European white majority is the basis of data collections. Unlike countries such as Canada and New Zealand (NZ), Australia has been slow in recognizing its ethnic diversity, not only socially but also through data collection. National censuses in Canada and NZ offer a good example of how these nations have reckoned with their ethnic diversity (Stevens et al. 2015; Cormack 2010). These two countries include questions on ethnic identity in their censuses, which include more comprehensive questioning and response options to self-identify. In doing so, the Canadian and NZ censuses better meet the international recommendations for collecting census information (United Nations 2017).

The five-yearly national census of population and housing in Australia, conducted by the Australian Bureau of Statistics (ABS), comprises the most comprehensive data collection in the country. The census has the potential to capture a more contemporary snapshot of the nation, so long as the questions asked are relevant and accurate (Howard 2019). Questions included in the census questionnaire must be sufficiently broad in their reach to be applicable and easily understood by all responding. In practice, this means census questions tend to be aimed at the majority: older Australians, who are on average less diverse in all manner of indicators than younger Australians (Allen 2020). Any questions proposed by the ABS for census fieldwork must also be agreed to by the Australian government (See, ABS NDb).

Census enumeration is intended to comprise active information giving by all individuals 15 years and over (ABS NDc). The act of responding to questions is akin to self-reporting, and thus serves as an empowering function of participation. Individuals are asked and thereby actively respond to the undertaking; none of the data items collected are done unto participants, for example imputed from tax records. Furthermore, the act of all people across Australia participating allows incomparable completeness and geographic coverage. In effect, hidden, hard to reach, minority, and disadvantaged population sub-groups are elevated to a level of inclusion that no other data collection enables. In other words, all people count.

2.2 DATA LANDSCAPE

The origin of data and the manner of its collection can increase the risks of data being more (or less) prone to embedded inequalities (Waring 2003). Whether data are collected as part of a total population enumeration (census), sample surveying, or are produced as administrative byproducts, methods of data collection can (by design) limit the opportunities for what can be seen or investigated (United Nations 2017). Representativeness — the degree to which characteristics and their distribution of the total population are reflected in data — is only one aspect of the capacity of data to manifest fairness. Other major considerations include whether data comprise active or passive collection strategies, and whether appropriate questions or variables are indeed asked or collated.

Australia is in an enviable position when it comes to social data (see e.g., the range of public data at: Australian Government, data.gov.au ND). The data landscape in Australia is vast, comprised of all manner of data collections, good availability of data for secondary analysis, survey reports, administrative byproducts, and community polls conducted by academic researchers, market researchers, representative bodies and government. Three notable Australian government-funded longitudinal surveys include the Housing, Income and Labour Dynamics in Australia study (HILDA) (Melbourne Institute ND), the Longitudinal Study of Australian Children

(LSAC) (Growing up in Australia ND), and the Longitudinal Study of Indigenous Children (LSIC) (Australian Government, Department of Social Services ND).

The expanse of data can falsely signal sufficient evidence to inform vital and pressing social issues concerning inequalities. For example, while HILDA and LSAC were based on nationally representative samples, overseas migrants (especially those newly migrated) and Australians experiencing social disadvantage are typically underrepresented among the participants (Watson 2011; Watson and Wooden 2004; Solof et al. 2005). This underrepresentation has the potential to adversely impact what analyses can be conducted using these data.

Government has traditionally been the gatekeeper of quantitative data and evidence. Government bodies fund data collection, research, and research functions (Howard 2019). In an attempt to increase knowledge about data holdings (especially those gathered through administrative processes) and promote wider transparency through accessibility, the national data commissioner and national data advisory council were established in 2018. The commissioner and council were tasked with “implementing a simpler and safer way to share and release public sector data to drive greater social and economic benefits for the community” (Australian Government, Office of the National Data Commissioner 2020a).

Furthermore, there are proposed changes to the national data environment, especially concerning potential improvements to access (Australian Government, Office of the National Data Commissioner 2020b; 2020c). Proposed legislation, in the Data Availability and Transparency bill, has the potential to make more government data accessible between government bodies and researchers alike. This proposed bill could pave the way for the end of the national population census and a move towards an increasing reliance on so-called big data (McIlroy 2020).

Regardless of the source of data, a data-driven initiative can empower individuals and communities (Waring 2003). A data-driven approach here refers to both

data to inform government policy and the efforts whereby data are specifically collected to reflect social trends and concerns. Lack of recognition through data can and does have real-world health impacts (Guest 2020). Identifying a social problem through measurement of its size, scope and impacts is just the start. While data is necessary for change, it alone is insufficient. Through a transformation, via analysis and reporting, data has the power to inform and initiate change. Data infrastructure thus supports or hinders the potential for research.

3. BETTER CAPTURING ETHNIC IDENTITY TO EMPOWER

Despite progress, sufficient or appropriate data on contemporary ethnic identity is not collected in Australia at a population level (Federation of Ethnic Communities Council of Australia 2020). Ethnic identity is a self-defined concept and can be a complex articulation of an individual’s biological background and social experiences and practices. This is especially the case as it relates to migrant and mixed-race populations (Green et al 2015). The inability of national data to reflect contemporary ethnic identity is not surprising given that census collection fails to recognize the way cultural identity is performed. For example, an Asian Australian might describe themselves as such. But across all the surrogate measures of ethnicity currently included in census, none of them can capture the ethnic identity of such a person in the data who is born in Australia to Australian-born parents and who does not speak a language other than English. In other words, the questions included in census to gather surrogate and indirect measures of ethnic identity — country of birth, language, religion and ancestry — are insufficient to reflect Australia’s diversity and cultural heritage (Khoo 1991; Perkins 2001). It is likely that Australia is far more diverse than the statistics signal.

Of particular importance is the inclusion in census of the ancestry question. A single question which simply asks, what is the person’s ancestry?, and only allows the identification of two options (ABS 2020). Such a question framing is reductionist and constrains ethnicity to

a narrow focus of biological race (Lepenies 2016). Notably, the question of ancestry is separate to that of Indigenous status.

Australia is lagging behind places like NZ, the United Kingdom (UK), and Canada on the collection of ethnic identity at a population level (Allan 2001; Stevens et al. 2015; Rocha et al 2018). This is a disappointing comparison given the ABS has, over time, been encouraged by advocates, researchers and government bodies to include a more comprehensive indicator of ethnicity and has failed to do so (ABS 2002; 1994). For example, in the Canadian census, respondents are asked about their ethnic identity and are able to choose from a list of over 500 options. Canadians can list multiple responses on the country's census, also enabling the collection of mixed ethnicity (Statistics Canada 2020). These data in the Canadian and NZ censuses also accommodate the interconnection with indigenous population identities.

Without the inclusion of a comprehensive question in census of ethnic identity among Australians, it is difficult (if not impossible) to truly gauge ethnic diversity, especially as it relates to how people view their own selves and expressions of culture. Sample surveys could not achieve an adequate understanding across the population due to the unknowns relating to the distribution of ethnic identity across population subgroups. Similarly, administrative data would have to be collected for all Australians and would thus be better from census. With traditional census-taking possibly being replaced by an integration of administrative data, the urgency of incorporating comprehensive questions in order to reflect contemporary ethnic identity in Australia is more important than ever.

A deficit in ethnic identity data impacts on the reflection of diversity in all manner of business and community life, adversely impacting the social capital and engagement of ethnically diverse people. It also has the potential to risk public health in times of a crisis, like the COVID-19 global pandemic. The measure of ethnicity, or lack thereof, acts to entrench inequalities because it instills an outdated notion of the ethnic composition of the nation as a majority white population with tokenistic nods to diversity.

Analysis is central to transforming data from mere numbers to achieve meaningful insights into social phenomena. From the variables used in statistical models to researchers themselves, all aspects of the analysis process when it comes to data and research have the potential for a continuance of the same. Researcher perspectives and previous literature lead to the building of statistical models and analysis to examine the world around us, but such bias can actually lead to the exclusion of the analysis of hidden inequalities. In the policy space, government economic modelling is used to forecast impacts of funding changes, among other things (Dee 1994). Similarly, economic modelling is used to inform government policy and practice. Much of the modelling by government, for example to explore the impacts on ethnically diverse people, focuses instead on those deemed as such by incomplete and inaccurate definitions (McLachlan et al. 2013).

Data exists within and emerges from complex and disparate legislation. The framework of infrastructure is outdated, failing to reflect contemporary Australia (Australian Government, Office of the Data Commissioner 2020a). Given this, it is no wonder that the data environment further acts to promote social divisions. Indeed, data holdings and the environment within which it comes should be built with investigations of inequality central. There is probably a no more apparent example of this than the matter of gender.

4. CONCLUSION AND DISCUSSION

Data and related processes can empower individuals and communities. In the least, data and related infrastructure should not create nor perpetuate unfairness and inequalities, especially in overlooking diversity. The collection and analysis of data and data infrastructure are opportunities for embedding equality and fairness principles. Introducing equality and fairness principles within data, data analysis, and infrastructure has the potential to inform and enable a change process to redress inequalities in more substantive and transparent ways. There are many opportunities which present themselves, especially in the context of transformation in the government data landscape, which could be effectively

leveraged to advance a program to promoting equality in and through data.

Examination of the collection of ethnic identity in Australia demonstrates the potential means by which data can constrain and adversely impact the potential for progress and change. While data alone cannot facilitate change, it is necessary for amendments to be made to the way the government collects, maintains and enables the use of its data to enable the redress of inequalities in contemporary Australian society. Ultimately, if Australians are reflected in data, they can be represented in society's social functions and decision making. Through representation comes equality and empowerment.

5. REFERENCES

1. Allan, J. 2001. "Review of the Measurement of Ethnicity: Classifications and Issues." Statistics New Zealand, September.
<https://www.stats.govt.nz/assets/Uploads/Retirement-of-archive-website-project-files/Methods/Review-of-the-Measurement-of-Ethnicity-Classifications-and-issues/review-of-the-measurement-of-ethnicity-classification-and-issues-main-paper.pdf> Accessed 30/04/2021.
2. Allen, L. 2020. *The Future of Us: Demography Gets a Makeover*. Sydney, Australia: NewSouth Publishing.
3. Australian Bureau of Statistics. 1994. "Testing of Ethnic Origin Questions for the 1996 Census" ABS Working Paper 94/4 — Ancestry, 1996, 30 November.
<https://www.abs.gov.au/ausstats/abs@.nsf/7d12b0f6763c78caca257061001cc588/a6845d677bc8c50fca2570610032c06d!OpenDocument> Accessed 11/2/2021.
4. Australian Bureau of Statistics. 2002. "Ancestry" 2001 Census of Population and Housing — Fact Sheet: Ancestry, 2001. 3 June.
<https://www.abs.gov.au/AUSSTATS/abs@.nsf/Lookup/2970.0.55.006Main+Features12001?OpenDocument>
5. Australian Bureau of Statistics. 2020. *2021 Census topics and data release plan*. 16 November.
<https://www.abs.gov.au/statistics/research/2021-census-topics-and-data-release-plan#data-download> Accessed 11/2/2021.
6. Australian Bureau of Statistics. NDa. *Multi-Agency Data Integration Project (MADIP) Research Projects*.
<https://www.abs.gov.au/websitedbs/D3310114.nsf/home/Statistical+Data+Integration+-+MADIP+Research+Projects> Accessed 10/05/2021.
7. Australian Bureau of Statistics. NDb. *ABS Media Statement on sexual orientation and gender identity questions and the 2021 Census*.
<https://www.abs.gov.au/websitedbs/d3310114.nsf/home/ABS%20Media%20Statement%20on%20sexual%20orientation%20and%20gender%20identity%20questions%20and%20the%202021%20Census> Accessed 28/02/2021.
8. Australian Bureau of Statistics. NDc. *Census*.
<https://www.abs.gov.au/census> Accessed 5/11/2020.
9. Australian Government, data.gov.au. ND.
<https://data.gov.au/search> Accessed 08/05/2021.
10. Australian Government, Department of Social Services. ND. *Footprints in Time: The Longitudinal Study of Indigenous Children*.
<https://www.dss.gov.au/about-the-department/publications-articles/research-publications/longitudinal-data-initiatives/footprints-in-time-the-longitudinal-study-of-indigenous-children-lsic> Accessed 08/05/2021.
11. Australian Government, Office of the Data Commissioner. 2020a. *Office of the Data Commissioner: Background*. Commonwealth of Australia, Canberra.
<https://www.datacommissioner.gov.au/about/background> Accessed 05/11/2020.

12. Australian Government, Office of the National Data Commissioner. 2020b. "Data Availability and Transparency Bill 2020: Exposure Draft." Department of the Prime Minister and Cabinet, Office of the National Data Commissioner, Consultation Paper September 2020. https://www.datacommissioner.gov.au/sites/default/files/2020-09/DAT%20Bill%202020%20exposure%20draft%20Consultation%20Paper%20Final_0.pdf Accessed 05/11/2020.
13. Australian Government, Office of the National Data Commissioner. 2020c. "Accreditation Framework." Department of the Prime Minister and Cabinet, Office of the National Data Commissioner, Discussion Paper. <https://www.datacommissioner.gov.au/sites/default/files/2020-09/Accreditation%20Framework%20Discussion%20Paper.pdf> Accessed 05/11/2020.
14. Cormack, D. 2010. *The Practice and Politics of Counting: Ethnicity Data in Official Statistics in Aotearoa/New Zealand*. Wellington: Te Rōpū Rangahau Hauora a Eru Pōmare. <https://www.otago.ac.nz/wellington/otago600095.pdf> Accessed on 10 February 2021.
15. Dee, P. 1994. "General Equilibrium Models and Policy Advice in Australia." Industry Commission Staff Information Paper: presented at IFAC Workshop on Computing in Economics and Finance, Amsterdam, 8–10 June. <https://www.pc.gov.au/research/supporting/general-equilibrium-models/generalequilibrium.pdf> Accessed 8/05/2021.
16. Federation of Ethnic Communities' Council of Australia, 2020. "If We Don't Count It...It Doesn't Count: Towards a Consistent National Data Collection and Reporting on Cultural, Ethnic and Linguistic Diversity". <https://fecca.org.au/wp-content/uploads/2020/10/CALD-DATA-ISSUES-PAPER-FINAL2.pdf> Accessed on 11 February 2021.
17. Green, E., Sarasin, O., and Fasel, N. 2015. "Immigration: social psychology aspects." In J.D. Wright (ed) *International Encyclopedia of the Social & Behavioural Sciences*, 2nd edition, volume 11. London: Elsevier.
18. Growing up in Australia. ND. *The Longitudinal Study of Australian Children*. <https://growingupinaustralia.gov.au> Accessed 08/05/2021.
19. Guest, A. 2020. "Mental health of LGBTIQ four times worse than general population." AM, Australian Broadcasting Corporation, 13 November. <https://www.abc.net.au/radio/programs/am/mental-health-of-lgbtqi-four-times-worse-than-general-population/12879872> Accessed 17 May 2021.
20. Horn, R. 1987. "Ethnic origin in the Australian census." *Journal of the Australian Population Association*, 4(1): 1–12.
21. Howard, C. 2019. "The politics of numbers: explaining recent challenges at the Australian Bureau of Statistics". *Australian Journal of Political Science*, 54:1, 65–81.
22. Khoo, S-E. 1991. "Consistency of ancestry reporting between parents and children in the 1986 census." *Journal of the Australian Population Association*, 8(2): 129–139.
23. Lepenies, P. 2016. *The Power of a Single Number: A Political History of GDP*. New York, US: Columbia University Press.
24. McIlory, T. 2020. "2021 Census could be Australia's Last Five-Yearly Population Snapshot." *Australian Financial Review*, 8 December. <https://www.afr.com/politics/federal/2021-census-could-be-australia-s-last-five-yearly-population-snapshot-20201207-p56l6n> Accessed 05/11/2020.
25. McLachlan, R., Gilfillan, G., and Gordon, J. 2013. "Deep and Persistent Disadvantage in Australia." Productivity Commission Staff Working Paper, July. <https://www.pc.gov.au/research/supporting/deep-persistent-disadvantage/deep-persistent-disadvantage.pdf> Accessed 8/05/2021.
26. Melbourne Institute. ND. *HILDA Survey*. <https://melbourneinstitute.unimelb.edu.au/hilda> Accessed 08/05/2021.

27. Morrissey, L. 2017. "Alternative Facts do Exist: Beliefs, Lies, and Politics." *The Conversation*. 5 October.
<https://theconversation.com/alternative-facts-do-exist-beliefs-lies-and-politics-84692>
Accessed 19/11/2020.
28. Nash, E. 2020. "Science, misinformation and dissent." *The Philosopher's Zone*, Australian Broadcasting Corporation, 22 November.
<https://www.abc.net.au/radionational/programs/philosopherszone/science-misinformation-and-dissent/12901888>
29. Perkins, M. 2001. "Australian mixed race." *European Journal of Cultural Studies*, 7(2): 177–199.
30. Rocha, L., Fozdar, F., Acedera, K., and Yeoh, B. 2018. "Mixing race, nation, and ethnicity in Asia and Australasia." *Social Identities* 25(3): 289–293.
31. Solof, C., Lawrence, D., and Johnstone, R. 2005. "Sample design." LSAC Technical Paper No. 1. Australian Institute of Family Studies, May.
<https://growingupinaustralia.gov.au/sites/default/files/tp1.pdf> Accessed 17 May 2021.
32. Statistics Canada. 2020. "Ethnic or cultural origins: Technical report on changes for the 2021 Census" 20 July.
<https://www12.statcan.gc.ca/census-recensement/2021/ref/98-20-0002/982000022020001-eng.cfm#a7>
Accessed 11 February 2021.
33. Stevens, G., Ishizawa, H., and Grbic, D. 2015. "Measuring race and ethnicity in the censuses of Australia, Canada, and the United States: Parallels and paradoxes." *Canadian Studies in Population*, 42(1–2): 13–34.
34. Timperio, A., Salmon, J., Telford, A., and Crawford, D. 2005. "Perceptions of local neighbourhood environments and their relationship to childhood overweight and obesity." *International Journal of Obesity*, 29: 170–175.
35. United Nations. 2017. "Principles and Recommendations for Population and Housing Censuses: Revision 3". Department of Economic and Social Affairs Statistics Division.
https://unstats.un.org/unsd/demographic-social/Standards-and-Methods/files/Principles-and-Recommendations/Population-and-Housing-Censuses/Series_M67rev3-E.pdf
Accessed 17 May 2021.
36. Waring, M. 2003. "Counting for something! Recognising women's contribution to the global economy through alternative accounting systems." *Gender and Development*, 11:1, 35–43.
37. Watson, N. 2011. "Methodology for the HILDA top-up sample". HILDA Project Technical Paper Series, No. 1/11, September.
<https://melbourneinstitute.unimelb.edu.au/assets/documents/hilda-bibliography/hilda-technical-papers/htec111.pdf> Accessed 17 May 2021.
38. Watson, N., and Wooden, M. 2004. "Sample attrition in the HILDA survey". *Australian Journal of Labour Economics*, 7 (3): 293–308.
39. Yu, M. 2017. "'I Am Not Your Negro' Gives James Baldwin's Words New Relevance." *All Things Considered*, NPR, Washington, 3 February.
<https://www.npr.org/2017/02/03/513311359/i-am-not-your-negro-gives-james-baldwins-words-new-relevance>