

What's in a Name?: A Cross-Section of Biography, Gender & Metadata in the *Design & Art Australia Online Database*

Tyne Daile Sumner
University of Melbourne
tdsumner@unimelb.edu.au,

Rachel Fensham
University of Melbourne
rachel.fensham@unimelb.edu.au

Nat Cutter
University of Melbourne
nat.cutter@unimelb.edu.au

Joanna Mendelsohn (DAAO)
University of Melbourne
joanna.mendelsohn@unimelb.edu.au

Abstract

In this paper, we present preliminary findings about issues identified by the Australian Cultural Data Engine (ACD-Engine) related to name and gender metadata practices, with a focus on the *Design & Art Australia Online* database (DAAO), and implications that extend to Australian cultural databases more broadly. Grounded in the history, context and specifics of data entry and management, we identify specific naming conventions, metadata contradictions and distinct institutional vestiges in the recording and representation of artists' careers. Using theoretical and statistical approaches, we categorize the types of variant names, showing a marked (but expected) contrast between men, women and non-binary individuals. By examining the affordances and constraints of naming conventions, we give attention to evolving trends in Australian data collection and use as they relate to the lives of individual artists. We argue that this local-level analysis is potentially applicable to wider transnational debates in humanities research and propose some new conceptual and technical approaches to the collection and use of biographical metadata in cultural databases.ⁱ

Keywords: name, metadata, gender, biography, arts, Australia, cultural data, digital humanities

1. Australian Cultural Databases: Background & Context

The Australian Cultural Data Engine (ACD-Engine) is an Australian Research Council (ARC) funded project aimed at improving the usability and interoperability of Australian cultural data. The ACD-Engine is interrogating the affordances of cultural data across a number of significant Australian cultural databases and thematic, including the relationship between biography and gender in artistic careers. Several significant cultural databases exist in Australia that are pertinent to gender research and women's history, including the *Australian Live Performance Database* (AusStage), the National Library of Australia's research portal Trove, *The Australian Women's Register* and the *Australian Dictionary of Biography*. In this paper we focus on one example that

provides a cross-section of the visual arts in Australia and facilitates comparative insights into the naming and gender histories of women artists: *Design and Art Australia Online* (DAAO).

The DAAO was established in 2007 as *The Dictionary of Australian Artists Online* as the result of efforts to provide a digital access point to the biographical works of art historians Joan Kerr and Vivien Johnson. Created by a consortium of universities, funded by five ARC grants and expanded by many independent contributors, the DAAO links peer-reviewed biographical data about Australian artists, designers, craftspeople and curators with information about artworks, event histories and art collections. Renamed *Design and Art Australia Online* in 2009, the DAAO is now one of the world's most comprehensive historical databases on Australian artists' careers and works, containing over 17,000 biographical records of varying length and detail.ⁱⁱ

2. Problem Statement

The cultural databases that the ACD-Engine has examined present challenges and new use cases for metadata conventions for name and gender. Dublin Core's current metadata terms do not adequately reflect and account for the name variations in existing cultural databases. Within cultural metadata structures we have identified that name variations, name changes and alternative names are inconsistently applied. In addition, cultural databases need to be responsive to evolving cultural and social understandings of gender. Contemporary research using cultural data, especially on questions of gender, increasingly depends upon access to metadata that appropriately records and reflects the relationships between biography, name and gender variations.

3. Methods and Sources

Examining the backend metadata and information architecture of the DAAO and its public-facing content reveals a range of ontological challenges surrounding naming conventions as they relate to the study and research of women's and non-binary people's lives and careers. As Tim Sherratt has suggested, humanities researchers should treat collection interfaces as archaeological sites, 'digging down through layers of technology, descriptive practice, and institutional history, to understand what is delivered so conveniently through our browsers'.ⁱⁱⁱ In each database, the institutional collecting histories (usually episodic) and data entry affordances (often voluntary and fragmented) make apparent the limitations of record keeping, while at the same time manifesting the distinctive characteristics of arts and cultural data as a trend. Dates are always given, but not always reliable; organisations and events are identified, however very often precise locations are missing; many stub entries exist, while descriptive content varies greatly in depth and quality. Search and retrieval tools are often contingent upon fixed, now long-operating technologies, and interoperability within the databases or beyond them to other resources can be variable. Most notably, the assignment of gender variables has been done manually by academic and/or volunteer contributors, and unique entries have rarely, if ever, been cross-checked by the subjects themselves.

Nevertheless, the thousands of entries in the DAAO, developed over nearly three decades, are powerful resources for the understanding of the macro- and micro-forces shaping cultural production and individual relationships in Australian art history. They are also a remarkable example of the cataloguing of 'names' that represent political, institutional and social patterns and conventions over time. Examining this collection closely exposes the influence that naming variations have had in fragmenting and shaping gender identities and the extent to which data entry protocols and database schemas have in turn shaped those identities and the formation of professional, popular and amateur careers.

We have undertaken a large-scale investigation of name changes in DAAO database. Recognising the incomplete nature of many of the more than 17,000 biographical entries, and the inconsistencies that arise examining data over more than 200 years of Australian history, we extracted a subset of 2,188 biographies containing all or most of certain key variables (name, gender, year of birth, biography or summary, artistic role, and career start date). Of these 2,188 individuals, 841 had at least one alternative name listed. The total number of alternative names across all individuals is 1,206. A simple gender breakdown of these 841 people – 465 men with

1,139 names (2.45/person), 373 women with 900 names (2.41/person), leaving aside for the moment the statistically insignificant three ‘unknowns’ with eight names between them (1.67/person) – would seem to suggest that men’s names are more variable. However, dividing the alternative names into distinct categories tells a rather different story, as we demonstrate below.

The guidelines on *Representing People’s Names in Dublin Core* (last updated in February 1998) list examples of how names ‘may be written down in alternative ways,’ including initials, nicknames and the use of middle names as given names, and highlights the importance for searching and sorting to list as many alternative name renderings as possible. It also provides an extensive and useful appendix on the range of naming conventions across cultures and languages, including name order, the use of patronymics and matronymics, the inclusion of prepositions (e.g. bin, de, von) and the use of religious names (e.g. Singh and Kaur for Sikhs).^{iv} However, there is no suggested schema for disaggregating the various kinds of alternative names or name changes that follow through individual lives, and so we have developed our own. By disaggregating name changes into five categories (listed in TABLE 1), we can see how alternative names affect men differently to how they affect women, making research on women’s lives more complex and challenging than those of men.

4. Findings

FIG. 1 below showcases the categories that we have devised. The category of Alternative Renderings comprises most alternative names (50.37%, n=269, for females and 80.36%, n=536, for males). However, we have found that women are substantially less likely to have Alternative Renderings listed and are thirty-one times more likely to have a Name Change listed, making them more than twice as likely to appear under a name substantially different from the ‘primary’. This issue alone means that women’s names present a substantially greater challenge to tracing and identifying the presence of women in historical records. One notable category of alternative names in the DAAO is Indigenous names. The high prevalence of and differences between male and female uses of Indigenous names likely emerges from the DAAO’s scholarly foundation in Vivien Johnson’s work on Indigenous artists, particularly women, and its ongoing support for coverage of Indigenous artistic traditions.^v By comparison, the differences in Pseudonym perhaps reflect the widely-recognised phenomenon that sees men receive nicknames (often in common use) more often than women, as opposed to indicating a lower prevalence of women adopting masculine pseudonyms (in Australian literature, perhaps most famous being Ethel Florence Lindesay Richardson, who published all her novels under the name ‘Henry Handel Richardson’).^{vi}

The categorizations the ACD-Engine have applied are based on readings of extant data, comparing alternative names to primary and other alternative names, rather than devised through more traditional research into individual artists’ biographies. One potential shortcoming in this approach is that it obscures when women used their initials instead of full given name/s as a gender-neutral ‘pseudonym’ (one might think of J. K. Rowling or P. D. James). Ascribing motivation to name changes presents ambiguity it is not made explicit in the textual data. Therefore, we have categorised all instances of using primary name initials as Alternative Renderings.

The statistical account we outline here substantially reflects the complexities of name changes and alternative names that have shaped women’s histories at the qualitative and aggregate levels discussed above. As databases progress and data structures such as Dublin Core engage in restructuring name entities, perhaps the question of how names enter databases in the first instance will need to progress a range of alternative mechanisms for tracking change.

5. Implications for Queer data and name

Data about people who identify as Lesbian, Gay, Bisexual, Transgender, Intersex, Queer and/or Questioning (LGBTIQ+) exists in distinct tension with data about names. In the broadest sense, queer data disrupt and challenge the datafication of gender binaries such as male/female or heterosexual/homosexual. In a more complex sense, queer data also bring about what Kevin Guyan describes as the ‘queering of data’ where the foundations upon which data categories stand, and

'the value granted to some identities above others and who *actually* benefits from the collection,' are critically questioned in the process of data collection, curation and use.^{vii}

As Guyan and others have shown, tensions and contestation over the design of classification systems for capturing data about gender are not new. Nor are the mechanisms of colonial power that continue to shape racial identity classifications that have direct implications for Indigenous communities and people of colour. In *Data Feminism* (2020), for instance, Catherine D'Ignazio and Lauren F. Klein show how 'standard practices in data science' tend to reinforce 'existing inequalities' that date back centuries, in particular those that benefit 'disproportionately elite, straight, white, able-bodied, cisgender men from the Global North'.^{viii}

Since some of the entries of people in the DAAO are those of living people, it is necessary to be attentive to the sensitivities of people who have transitioned genders or are now non-binary. The DAAO has four fields for assigning gender: male, female, unspecified and unknown. The assigning of these categories to individuals in the database can be changed by the moderators. There is also a moveable hierarchy of names, which means that it is possible to privilege an artist's new name, while not denying past achievements under a previous name and/or gender. For example, the artist now known as 'D Harding' has 'Dale Harding' entered as an alternative name, and their gender is now 'unspecified.' This metadata classification operates in a somewhat different manner to the data about the Netflix streaming original series *Umbrella Academy*, in which Ellen Page has been retrospectively changed to Elliot Page in all three seasons of the series' online record.

As these examples show, data categories that may seem objective and merely bureaucratic to the average person are often highly subjectivised and produced under conditions of bias and self or social preservation. Moreover, as Lauren E. Bridges explains, 'entropic, fugitive, and queer data', histories of categorization and naming 'have long been entangled in histories of sovereignty, colonialism, subjugation and exploitation'.^{ix} Reflecting this reality, Bowker and Star rightly argue:

We have a moral and ethical agenda in our querying of these systems. Each standard and each category valorizes some point of view and silences another. This is not inherently a bad thing—indeed it is inescapable. But it *is* an ethical choice, and as such it is dangerous—not bad, but dangerous.^x

Querying and updating the classification systems for gender and name in cultural databases therefore requires a synthesis of intellectual accountability, reflecting culturally and socially informed work *with* technical expertise. When it comes to name data in cultural databases, it is therefore encouraging that metadata standards and taxonomies are slowly evolving to reflect the fluidity of gender and the need for more diverse understandings of gender. One example is the Australian Government's 'Standard for Sex, Gender, Variations of Sex Characteristics and Sexual Orientation Variables,' developed by the Australian Bureau of Statistics (ABS) to 'standardise the collection and dissemination of data relating to sex, gender, variations of sex characteristics and sexual orientation.'^{xi} The 2020 standard 'for each variable includes the concept(s), definition(s), questionnaire modules, classification, coding structure, and output categories to be used in ABS interviewer-based and self-enumerated collections. It also provides 'guidance on deriving cisgender (cis) and trans and gender diverse (trans) counts using the sex and gender variables.'

Outside Australia, another example is the American Library Association's *Gay, Lesbian, Bisexual and Transgender (GLBT) Controlled Vocabularies and Classification Schemes*, first compiled by Matt Johnson in 2007 and updated by Jessica L. Colbert in 2017.^{xii} As these histories of databases and vocabulary schemas suggest, the connection between name and gender is both temporal in its dimensions and reflects the extent to which cultural databases are, at their core, human driven and must reflect the changing complexities of human identification and subjectification. A name in a cultural database thus becomes a representative – if not the exemplary – instance of the interplay between one's personal perception of themselves and their public persona in the arts and cultural sphere. While naming can be performative in practice, it can also be revealed as a sign of power in the process of datafication.

6. Recommendations

Based on our comparative analysis of naming conventions in Australian cultural databases (especially the DAAO) with the naming metadata used in the current iteration of Dublin Core, we suggest an incremental approach for updates to metadata entry of names that would facilitate a more robust and responsive understanding of name diversity and name changes, given the broad implications for tracing women's lives through and across cultural data. We suggest a division of alternative names between, at a minimum, name changes that replaced earlier names (including those related to marriage and legal name changes), names adopted alongside legal names for public purposes (pseudonyms, stage names, 'local' names for migrants, digital handles, honorifics, brand names and usernames), and alternative spellings and initials (including, potentially, variations across languages in historiography, or wrought by migration). Each of these categories offers distinctive information and deserves more granular consideration.

7. Figures and Tables

Alternative Name Category	Category Definition	% Female Alt. Names	% Male Alt. Names
Alternative Rendering	Close variations on the primary name, e.g. the addition or exclusion of middle names, various kinds of initials, contracted or common nicknames e.g. Chris for Christopher, Jack for John, alternative phonetic spellings, and altered name order.	50.37% (n=269)	80.36% (n=536)
Migration Alternative Rendering	Chiefly Anglicised names with similar meanings or orthographies to the primary name	2.06% (n=11)	1.95% (n=13)
Name Changes	Legal changes e.g. at marriage or by deed poll, chiefly changes to surname	31.46% (n=168)	1.50% (n=10)
Indigenous Names	Chiefly skin names or names of Indigenous origin used without any other names and distinct from surnames or patronymics	7.49% (n=40)	3.90% (n=26)
Pseudonyms	Stage names, pen-names, and nicknames that are dissimilar to the primary name	8.61% (n=40)	12.29% (n=82)

TABLE 1. Categories of alternative names as listed in the DAAO. The categories are of the author's devising, attempting to produce an appropriate schema for disaggregating and comparing the various kinds of alternative names that appear in the dataset, that would reveal gendered patterns of use.

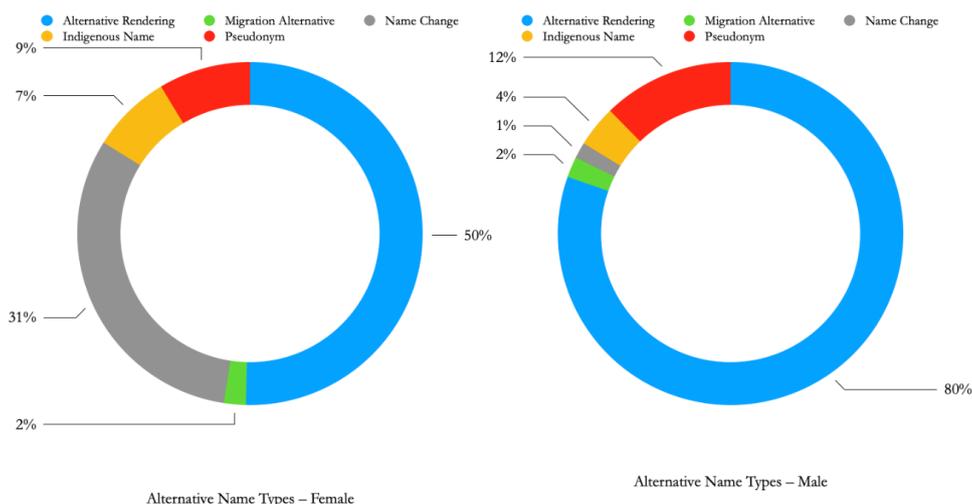


FIG. 1. Graphs indicating the relative prevalence of alternative name categories in a subset of DAAO biographies. Generated by the authors from open-source data.

Acknowledgements

This paper is supported by the Australian Research Council LIEF project, The Australian Cultural Data Engine for Research, Industry and Government (<https://www.acd-engine.org/>)

References

- AusStage, 'About AusStage', accessed 6 May 2022, <https://www.ausstage.edu.au/pages/learn/about/>
- Jonathan Bollen, Neal Harvey, Julie Holledge, and Glen McGillivray, 'AusStage: e-Research in the Performing Arts', *Australasian Drama Studies* 54 (2009), 178-94
- Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass: MIT Press, 1999)
- Catherine D'Ignazio, and Lauren F. Klein, *Data Feminism* (Cambridge, Mass.: MIT Press, 2020)
- Kevin Guyan, *Queer Data: Using Gender, Sex and Sexuality Data for Action* (London: Bloomsbury, 2022)
- Matt Johnson, 'GLBT Controlled Vocabularies and Classification Schemes', *Rainbow: a Round Table of the American Library Association*, August 2007, accessed 6 May 2022
- Vivien Johnson, *Aboriginal Artists of the Western Desert: a biographical dictionary* (Roseville East, NSW: Craftsman House, 1994).
- Tim Sherratt, 'Hacking Heritage: Understanding the Limits of Online Access,' in *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*, edited by H. Lewi et al. (London: Routledge, 2019).

ⁱ The authors would like to thank Scott East for his feedback on and engagement with an earlier draft of this paper.

Footnotes

ⁱⁱ AusStage, a comparable Australian cultural database, was created in 1999 as a collaboration between theatre scholars, government, and industry to address the need for research information in Australian theatre, drama and performance studies by building an index of performing arts events in Australia, and a directory of research resources on the performing arts. As of mid-2022, AusStage has entries on over 125,000 events, 185,000 contributors, 20,000 organisations and 12,000 venues from the late eighteenth century to today. See 'About AusStage', AusStage, accessed 6 May 2022,

<https://www.ausstage.edu.au/pages/learn/about/>; Jonathan Bollen, Neal Harvey, Julie Holledge, and Glen McGillivray, 'AusStage: e-Research in the Performing Arts', *Australasian Drama Studies* 54 (2009), pp. 178-94.

ⁱⁱⁱ Tim Sherratt, "Hacking Heritage: Understanding the Limits of Online Access," in *The Routledge International Handbook of New Digital Practices in Galleries, Libraries, Archives, Museums and Heritage Sites*, ed. H. Lewi et al. (Routledge, 2019).

^{iv} Andrew Waugh, 'Representing Peoples Names in Dublin Core', published 3 February 1998, accessed 5 August 2022, <https://www.dublincore.org/specifications/dublin-core/name-representation/>.

^v See Vivien Johnson, *Aboriginal Artists of the Western Desert: a biographical dictionary* (Roseville East, NSW: Craftsman House, 1994). The DAAO's listing of 1914 Indigenous Australian artists provides particular challenges in naming artists and their communities. As is common with First Nations people in other countries, Australia's Indigenous artists have experienced both physical removal and cultural dislocation at the hands of authorities. To address this legacy of colonization requires recognition that heritage country may not be the same as the artist's geographic location. The DAAO has fields to note language groups and, for those artists still closely connected to country, their dreaming. The DAAO recording of places and peoples also takes into account variations in spelling, a result of the cultural background of the colonisers' missionaries, who first wrote down the words which had been spoken for thousands of years. For example, in the 19th and 20th centuries the people of Central Australia have been recorded as "Aranda", "Arrunta", and "Arunta". A search of the DAAO using any of those variations, corrects to the preferred spelling, "Arrernte". This legacy of colonization is reflected in the DAAO's inclusion of the 650 biographies from Vivien Johnson's Storylines, an ARC research project that is now only otherwise accessible using the Wayback machine. See

<https://web.archive.org/web/20160205054037/http://www.storylines.org.au/>.

^{vi} See Vivian de Klerk and Barbara Bosch, 'Nicknames as Sex-Role Stereotypes', *Sex Roles* 35 (1996), p. 525-41.

^{vii} Kevin Guyan, *Queer Data: Using Gender, Sex and Sexuality Data for Action* (London: Bloomsbury, 2022), p. 1.

^{viii} Catherine D'Ignazio, and Lauren F. Klein, *Data Feminism* (Cambridge, Mass.: MIT Press, 2020), p. 8.

^{ix} Lauren E. Bridges, 'Digital Failure: Unbecoming the "Good" Data Subject through Entropic, Fugitive, and Queer Data', *Big Data & Society* 8, 1 (2021), p. 2.

^x Geoffrey C. Bowker and Susan Leigh Star, *Sorting Things Out: Classification and Its Consequences* (Cambridge, Mass: MIT Press, 1999), pp. 5-6.

^{xi} Australian Bureau of Statistics, 'Standard for Sex, Gender, Variations of Sex Characteristics and Sexual Orientation Variables', updated 14 January 2021, accessed 10 August 2022, [HYPERLINK "https://www.abs.gov.au/statistics/standards/standard-sex-gender-variations-sex-characteristics-and-sexual-orientation-variables/latest-release#references-and-resources."](https://www.abs.gov.au/statistics/standards/standard-sex-gender-variations-sex-characteristics-and-sexual-orientation-variables/latest-release#references-and-resources) [https://www.abs.gov.au/statistics/standards/standard-sex-gender-variations-sex-characteristics-and-sexual-orientation-variables/latest-release#references-and-resources.](https://www.abs.gov.au/statistics/standards/standard-sex-gender-variations-sex-characteristics-and-sexual-orientation-variables/latest-release#references-and-resources)

^{xii} The materials compiled are the result of decades of consultation and critique, which began with a Round Table at the ALA's general meeting in Dallas in 1971, in which the then Task Force on Gay Liberation criticized the homophobic categorisations and labelling that were used in the Library of Congress Subject Headings and the Dewey Decimal Classification system. Matt Johnson, 'GLBT Controlled Vocabularies and Classification Schemes', *Rainbow: a Round Table of the American Library Association*, August 2007, accessed 6 May 2022, <https://www.ala.org/rt/rt/popularresources/vocab>.