

Is a Picture Worth 10,175 Australian Novels?

[S]cience ... is rhetoric, a series of efforts to persuade relevant social actors that one's manufactured knowledge is a route to a desired form of very objective power. Such persuasions must take account of the structure of facts and artifacts, as well as of language-mediated actors in the knowledge game.¹

[D]ifferent types of data derived from different sources, which, interpreted cautiously ... can illuminate and explain processes within book history that are simply not visible by any other means.²

Embedded within any statistical analysis of Australian bibliographic data are definitional issues over the research sample which reflect some of the fundamental problems in thinking about the commodity-text (or book) in a singular, national context. If, as Amanda Petrucci claims, the bibliographic sciences display "a profound ideological bias, masked by a penchant for abstract, objective technology",³ the issue of how certain books are selected, appropriated and inherited by a group of readers as being meaningfully "Australian" becomes an important methodological challenge to any statistically-informed findings. This study uses the results from applying statistical analysis to two data sources: "AustLit, The Australian Literature Resource"⁴ and "Libraries Australia",⁵ the former providing "information on hundreds of thousands of creative and critical Australian literature works

¹ Donna Haraway, "Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective" *Feminist Studies* 14.3 (1988): 575-599, 577.

² Simon Eliot, "Very Necessary but Not Quite Sufficient: A Personal View of Quantitative Analysis in Book History", *Book History* 5 (2002): 283-293.

³ Amanda Petrucci, "Reading to Read: A Future for Reading", in Guglielmo Cavallo and Roger Chartier, eds, *A History of Reading in the West*, Cambridge: Polity Press (1999): 350.

⁴ <http://www.austlit.edu.au/>.

⁵ <http://www.nla.gov.au/librariesaustralia/>.

relating to more than 100,000 Australian authors and literary organisations”,⁶ and the latter comprising the Australian National Bibliographic Database (ANBD), the global catalogue WorldCat and the British Library catalogue. The question then driving this chapter is whether the creation of a data source is cultural work and whether this impacts the presentation of historiography interpreted from a data source — especially so with data that is centred on culture or nation as its core organising concept (even though intellectually we would argue that the nation-state is no longer obsessively maintained as a “categorical foundation or operational centre”).⁷ In a study that contributes to a revisionary account of the circulation of Australian culture in the world by joining quantitative analysis with documentary traces, responding to this question will add to future applications of new empiricism.

The use of numerical comparisons that computational analysis affords is not a problem-free exercise in Australian literary and print culture history. Since 2006 it has been possible to engage with AustLit tagged-text data along the lines of enquiry suggested by Moretti and William St Clair. In the absence of proprietary software suiting the research needs of this study, this has meant building functions using PHP (Hypertext Preprocessor)⁸ and MySQL (Structured Query Language)⁹ that enact specific analytical outcomes (see chapter three). These outcomes, presented within the context of “new empiricism” at ASAL¹⁰ and mini-ASAL conferences during 2007 and 2008, represent many hours of data mining, function programming and rendering. It may seem an odd choice of description for this work but the word “rendering” is deliberately used to suggest the practice of 3D computer graphic modelling; where an underlying mesh, in this case a vast resource of publication data legitimately downloaded from the AustLit website, is worked through a series of hand-made, hand-coded tools to generate useable representations for academic debate. As these representations do not wear their underlying design on their sleeve, the resulting images of statistical analysis, deployed in research for the purposes of discussing publication trends in Australian literary history, tend to elide their links with the technological labour that preceded their creation. In this sense, one might say — with apologies to Van Maanen who is writing about ethnography — that the “fieldworker, having finished the job of

⁶ “About AustLit”, <http://www.austlit.edu.au/about>.

⁷ Rita Raley, “eEmpires”, *Cultural Critique* 57 (Spring 2004): 126.

⁸ See “PHP: Hypertext Preprocessor,” online (2009), <<http://au.php.net/>>, accessed April 2009.

⁹ See “MySQL: The world’s most popular open source database,” online (2009), <<http://www.mysql.com/>>, accessed April 2009.

¹⁰ The Association for the Study of Australian Literature.

collecting data, simply vanished behind a steady descriptive narrative justified largely by the respectable image and ideology of ... [new empiricist] practice”.¹¹ Which is to say, in using computer technologies to facilitate interpretive work the statistical graphs placed “a premium on surface manipulation and thinking in ignorance of [their] underlying mechanism”.¹² Essentially, it asked viewers to suspend disbelief and become absorbed in, even seduced by, a “certain kind of secular magic” that was being performed on the screen.¹³ As Martyn Jessop claims, “Images are seductive and there is a natural tendency to instinctively believe whatever one sees with one’s own eyes but in the case of digital visualisations what is seen is entirely a constructed object”.¹⁴

This observation is important because “new empiricism” and its related practices capitalise on the notion of computers employing neutral, carefully structured logic with an absence of poetics and felt emotion. Indeed, it is the ways that computers “think” which is taken to be “their most culturally important characteristic”¹⁵ and contemporary social rhetoric surrounding technology encourages a view of computers as communicating (or “thinking”) in a logic that proceeds towards very specific ends. Neil Postman and Andrew Postman in their critique on the decline of the printed word have referred to this as the “cognitive biases and social effects” which follow the use of computers.¹⁶ New empiricism, in denoting precise rational procedures linked with computing, seeks to be an expression of those ends and is connected with the production of digitally-based visual texts that, like this study’s statistical graphs in chapter three, seemingly “speak for themselves” about Australian literary history. This might be because “the kind of knowledge the computer encourages is rationalist, linear and analytic, mimicking the public communication of science”¹⁷ and the possibility of objectivity, which the humanities it is claimed secretly desires.¹⁸

¹¹ J. Van Maanen, *Tales of the Fields*, Chicago: University of Chicago Press (1998): 46.

¹² S. Turkle, *Life on the Screen*, London: Weidenfeld and Nicolson (1996): 35.

¹³ Nigel Thrift, “The Material Practices of Glamour” *Journal of Cultural Economy* 1.1 (2008): 9-23, 9.

¹⁴ Martyn Jessop, “Digital Visualisation as a Scholarly Activity” *Literary and Linguistic Computing* 23.3 (2008): 290.

¹⁵ Julia Flanders, “Data and Wisdom: Electronic Editing and the Quantification of Knowledge” *Literary and Linguistic Computing* 24. 1 (2009): 55.

¹⁶ Neil Postman and Andrew Postman, *Amusing Ourselves to Death: Public Discourse in the Age of Show Business*, London: Penguin (2005), Kindle Edition, Locations: 2,390-401.

¹⁷ Lynette Hunter, *Critiques of Knowing: Situated Textualities in Science, Computing and the Arts*, New York: Routledge (1999): 86.

¹⁸ Martyn Jessop, “Digital Visualisation as a Scholarly Activity” *Literary and Linguistic Computing* 23.3 (2008): 282.

Information systems and information use are also highly “socio-technical in nature [...] ... they develop their own personality as determined through the initial design of the system and its ongoing human interface, and they reflect the politics of the organisational structure and its human actors”.¹⁹ Perhaps new empiricism, in its perceived relevance to Australian literature and the humanities in general, is a system of analysis that represents what Fredric Jameson lamented as the “depthlessness” of postmodernism,²⁰ privileging the consumption of visual images over deeper, critical forms of thinking? Indeed, does the move from “close” reading to “distant” reading parallel the loss of the felt authenticity of emotion and the rise of simulation and surface? Such questions are beyond the scope of this chapter. However, if changes in “technologies do not just expedite ... knowledge transmission, but deliver it in alternative ways which require different interpretive and behavioural skills”,²¹ then by considering the embodiment of the disciplinary space of Australian literature on a computer screen (through AustLit) as a type of “cultural work”, we might begin to take account of “the representational logic of the [computer] medium” in discussions of empiricism and modern-day forms of Australian literary knowledge production.²²

This chapter explores the work behind the charts and graphs presented in chapter three. This will include the necessary apologetics and methodological uncertainties that contextualise analytic labour, and it will put forward an alternative reading of new empiricism which suggests that internet and computing technologies are shaping the cultural grammar of the domain of Australian literature in ways yet to be fully understood but in ways which need to be corralled methodologically. It will propose that in the contemporary humanities environment new empiricism should continue to provide important “reference points from which qualitative data can be understood”²³ and as a way for literary scholars to visualise quantitative research but from within the framework of an *Australian Charter for the Computer-Based Representation of Literary History*. In so doing, this chapter will draw upon standards from *The London Charter*. Established in relation to Cultural History, *The London Charter* has argued that “computer-based visualisation

¹⁹ Robert Hutchinson, “Knowledge and Control: A Marxian Perspective on the Productivity Paradox of Information Technology” *Rethinking Marxism* 20.2 (April 2008): 288-304.

²⁰ Frederic Jameson, *Postmodernism or, the Cultural Logic of Late Capitalism*, Durham: Duke University Press (1991).

²¹ Margery Mayall, “Attached to their Style: Traders, Technical Analysis and Postsocial Relationships” *Journal of Sociology* 43.4 (2008): 421-437, 422.

²² Detlev Zwick and Nikhilesh Dholakia, “Bringing the Market to Life: Screen Aesthetics and the Epistemic Consumption Object” *Marketing Theory* 6 (2006): 41-62, 42.

²³ Fiona Black, Bertrum MacDonald and J. Malcolm Black, “Geographic Information Systems: A New Research Method for Book History” *Book History* 1. 1 (1998): 13.

methods” should be “applied with scholarly rigour, and that the outcomes of research that include computer-based visualisation should accurately convey to users the status of the knowledge that they represent, such as distinctions between evidence and hypothesis, and between different levels of probability”.²⁴ This is not to adjudicate what shape and form an *Australian Charter* might take but rather to raise the possibility of an in-built scholarly apparatus for empiricism in Australian literary history.²⁵ It should also be noted that *The London Charter* is not the only feasible template: the *Text Encoding Initiative* (for scholarly editors) and the *Electronic Cultural Atlas Initiative* are other possible models for standards.²⁶

Admittedly, the term “new empiricism” has been made to carry much rhetorical weight thus far and it needs further definition beyond its use here as an implied synonym for “book history” or for quantitative analysis in the humanities via computing (though it can be these things). To begin with, new empiricism is not “e-Research” nor “e-Literature” nor anything where the lower-case letter “e” continues to “operate as the value-added, universal signifier of the brave new wired world”.²⁷ It can, however, be linked with such projects, sometimes as an internet-hosted digital tool at the service of e-Research or e-Lit, other times as a particular mode of quantitative enquiry applied within the humanities to a dataset. The core attribute shared between both approaches is that “new empiricism” is — ideally — the theoretical position in Australian literary history and Australian print culture studies where information systems and information use merge with qualitative historiography in the discovery of new knowledge through data mining, data analysis and, often, digital visualisation.²⁸ (Representations or summaries of data as lists ordered according to a specific enumerative calculation, to graphs depicting various statistical correspondences, broadly indicate the kinds of combinations of data analysis and digital visualisation that can occur within the context of “new empiricism”.)

²⁴ “Preamble”, *The London Charter for the Computer-Based Visualisation of Cultural Heritage* (Draft 2.1), online (2009), <<http://www.londoncharter.org/>>, accessed April 2009.

²⁵ *The London Charter* recommends that “Each community of practice, whether academic, educational, curatorial or commercial, should develop London Charter Implementation Guidelines that cohere with its own aims, objectives and methods”. *Ibid.*, p. 5.

²⁶ I am grateful to Tim Dolin for bringing these to my attention: the *Text Encoding Initiative* for scholarly editors <<http://www.tei-c.org/index.xml>> and the *Electronic Cultural Atlas Initiative* <<http://www.ecai.org/>>.

²⁷ Rita Raley, “eEmpires”, *Cultural Critique* 57 (Spring 2004): 111.

²⁸ Digital visualisation focuses on “arranging, formatting, or otherwise transforming the appearance of data”. Willard McCarty, “Humanities Computing”, *Encyclopaedia of Library and Information Science* (2003): 1230.

In modern print culture studies, this is what Moretti refers to as “distant reading” or the “quantitative approach to literature”, in which a large collective system might be grasped as a whole through computing and graphical aids and where an individual text’s relationship to the whole may be charted. It is also what Martin Mueller facetiously refers to as “not reading”²⁹ in which, as Sculley and Pasanek equally claim, distant reading requires the researcher to “trade in a close reading of the original text for something that looks like a close reading of experimental results — a reading that must navigate ambiguity and contradiction”.³⁰ While Moretti’s and Mueller’s terms are valid, neither fully captures new empiricism. Though Moretti’s term is often an interchangeable referent for new empiricism, the word “distant” as an antonym to “close” implies “objectivity” and therefore capitalises on this imported association as being a less “intimate”, less “sentimental”, more scientific type of reading without actually explicitly claiming it is so. Similarly, Mueller’s “not reading” obscures the irony that where “we had hoped to explain or understand those larger structures within which an individual text has meaning in the first place, we find ourselves acting once again as interpreters”.³¹ That is, through analysing charts and graphs, we engage in the kind of literary criticism and literary reading practices which new empiricism supposedly distances itself from.³² It is what John Unsworth refers to as a process of “computation into criticism”.³³ Scholars need to take account of these stances, but new empiricism’s relative youthfulness in Australian humanities departments means that there is a “lack of in-built scholarly apparatus” such as an *Australian Charter* to provide an agreed theoretical position and methodological direction on what constitutes good information use and sound data visualisation.³⁴

²⁹ Tanya Clement, Sara Steger, John Unsworth and Kirsten Uszkalo, “How Not to Read Million Books”, online (2009), <<http://www3.isrl.illinois.edu/~unsworth/hownot2read.html>>, accessed April 2009.

³⁰ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 417.

³¹ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 417.

³² Quoted in D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 411.

³³ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 417.

³⁴ Martyn Jessop, “Digital Visualisation as a Scholarly Activity” *Literary and Linguistic Computing* 23.3 (2008): 288.

This motion for an *Australian Charter* is also intensified by a view that there can be something mildly suspect about new empiricism, in that its structure of representation can in some modes resemble economic rationalism whereby knowledge becomes most valuable when it is “quantifiable and hence offers comparability”.³⁵ Economic rationalism of course, as literary culture’s Other, is the “belief that everything of value can be considered in economic terms”.³⁶ Often, this has led to a “dangerous equivocation” for the bureau-/techno-cratic class that administrates the quantifiable research contributions of humanities departments in Australian universities — “namely, thinking that since any x can be described in (more or less metaphorically) informational terms, the nature of any x is genuinely informational”.³⁷ (As an analogy, consider the impact that the introduction of book sales data-monitoring software like Nielsen Bookscan³⁸ had on Australian literature, igniting debates — continuing today — that link literary fiction’s performance in the marketplace with questions targeting Australian literature’s continuing relevance in modern education.)³⁹

New empiricism is an emerging attractor in humanities scholarship and funding applications, reflecting the mathematical logic that is generally ascendant in advanced societies like Australia, and which marks a new theoretical position where modern literary research might usefully converge.⁴⁰ This chapter thus is cautious about informational methodologies but it is not exhaustive, as the “formal bias of socially rational artefacts and institutions is far more difficult to identify and criticize than inherited mythic and traditional legitimations”.⁴¹ However, in common with much screen-based analysis is the tendency to consider only the

³⁵ Roger Beckmann, “Literature, Science and Economic Rationalism”, in Katherine Barnes and Jan Lloyd-Jones, eds., *Words for their Own Sake: The Pursuit of Literature in an Economic Rationalist World*, Melbourne: Australian Scholarly Publishing (2004): 58-69, 63.

³⁶ Roger Beckmann, “Literature, Science and Economic Rationalism”, in Katherine Barnes and Jan Lloyd-Jones, eds., *Words for their Own Sake: The Pursuit of Literature in an Economic Rationalist World*, Melbourne: Australian Scholarly Publishing (2004): 63.

³⁷ Luciano Floridi, “What is the Philosophy of Information?”, *Metaphilosophy* 33. 1/2 (January 2002): 140.

³⁸ Nielsen Bookscan Australia, online (2009), <<http://www.nielsenbookscan.com.au/>>, accessed April 2009.

³⁹ See Rosemary Neill, “Who is Killing the Great Books of Australia? Lits Out”, *The Australian*, 18 March 2006; Rosemary Neill, “Is this the most feared man in Australian literature? - The Biblio Files”, *The Australian*, 22 July 2006.

⁴⁰ See, for example, Katherine Bode and Robert Dixon (eds), *Resourceful Reading: The New Empiricism, eResearch and Australian Literary Culture*, Sydney: Sydney University Press (2009).

⁴¹ Andrew Feenberg, “From Critical Theory of Technology to the Rational Critique of Rationality” *Social Epistemology* 22.1 (January-March 2008): 25.

screen's "output and its particular relevance to one's purposes"⁴² rather than interrogate the underlying systems — the "technological unconscious"⁴³ as it were — which insist on the potency of screen-based analysis. The argument I put forward then, drawing on sociological and cultural studies readings of technology, is that through new empiricism the "aesthetic qualities of a visual representation governed by the screen" is producing a "new ontology" of Australian literary history.⁴⁴ It is in this view that an *Australian Charter*, progressively configured by the academy within the disciplinary context of Australian literature and built upon the principles established by *The London Charter*, might enable the application of new empiricism to cohere more critically with the aims, objectives and methods of Australian literary history. This would help ensure scholars take full account of the ideological or methodological shifts still unfolding within the discipline through the contemporary turn towards computer-based — and internet-hosted — visualisation techniques.⁴⁵ Rather than centre on the question of how can scholars use new empiricism to enhance Australian literary studies, we might instead ask how can scholars use Australian literary studies to enhance new empiricism?

Necessary Apologetics

Applying methods of new empiricism to publication data drawn from the official Australian bibliographic record has been a particularly daunting process since 2006, complicated by issues of technology, problems of logic and limits to data. As a result, any published article about the application of new empiricism to Australian literary history has started with the necessary apologetics — "necessary" because researchers frequently work with materials that can be incomplete or unfinished and therefore it is considered good scholarly practice to situate the communication of any findings with disclaimers that address methodological gaps. This is so that others may verify new knowledges, oppose them or even build on them, things that seem the core activities of humanities research.

⁴² Karin Knorr Cetina, "Sociality with Objects: Social Relations in Postsocial Knowledge Societies" *Theory, Culture and Society* 14 (1997): 1-30, 8.

⁴³ Andrew Feenberg, "From Critical Theory of Technology to the Rational Critique of Rationality", *Social Epistemology* 22.1 (January-March 2008): 5-28, 18.

⁴⁴ Zwick and Dholakia, "Bringing the Market to Life: Screen Aesthetics and the Epistemic Consumption Object", p. 42.

⁴⁵ *The London Charter* recommends this activity for all subject areas. See 4.1, Richard Beacham, "An Introduction to the London Charter", *The E-volution of IC Technology in Cultural Heritage*, papers from the Joint Event CIPA / VAST / EG / EuroMed Event (2006): 4.

As recent examples of this, Toni Johnson-Woods in her 2008 *JASAL* article on the Carter Brown Mystery Series describes in a “data collection apologia” that “[t]he problems encountered during the course of ... [her] project”, in the creation of a complete Carter Brown bibliography, “are common in literary historiography”.⁴⁶ Priya Joshi in her analysis of the English novel in India notes that some data, pertinent to a sustained intellectual history of reading ... [remains] extremely elusive”.⁴⁷ Tim Dolin through the Australian Common Reader project⁴⁸ reveals an extraordinarily rich history of Australian reading based on the surviving loan records of seven community-based libraries,⁴⁹ but some data spans only eighteen months, prompting Roger Osborne to warn that “[g]eneral conclusions from this limited dataset must be cautious”.⁵⁰ Moreover, Carol Hetherington, in examining the American long-distance connection in Australian literature, has raised questions about the stability or incompleteness of book histories relying on legacy print-based bibliographic texts, materials which have traditionally been considered “impeccable ... resources”.⁵¹

Despite these gaps in the archival record, Dolin’s findings (along with Johnson-Woods, Joshi and Hetherington) challenge assumptions that have informed previous histories of reading, publishing and literature. Yet in what seems shared methodological territory or interpretative strategies, there are strong reservations about new empiricism and what might be characterised as an early or premature adoption of it in Australian literature, particularly in the application of quantitative forms to publication data. Central to these concerns is the empirically vast AustLit database⁵² as it progresses towards significant milestones but which also leaves some information (at the time of writing) during this crucial maturing phase “insufficiently comprehensive ... for statistical analyses”.⁵³ This is a

⁴⁶ Toni Johnson-Woods, “The Promiscuous Carter Brown” *Journal of the Association for the Study of Australian Literature*, online (2 May 2008): 167.

⁴⁷ Tim Dolin, “The Secret Reading Life of Us”, in Brian Matthew, ed., *Readers, Writers, Publishers: Essays and Poems*, Canberra: Australian Academy of the Humanities (2004): 117.

⁴⁸ Australian Common Reader project, online (2008), <<http://www.australiancommonreader.com>>, accessed April 2009.

⁴⁹ Tim Dolin, “The Secret Reading Life of Us”, in Brian Matthew, ed., *Readers, Writers, Publishers: Essays and Poems*, Canberra: Australian Academy of the Humanities (2004): 134.

⁵⁰ Roger Osbourne, “Australian Common Reader”, *Aus-e-Lit: Collaboration, Integration and Annotation*, online (12 September 2008), <<http://aus-e-lit.blogspot.com/2008/09/australian-common-reader.html>>, accessed April 2009.

⁵¹ Carol Hetherington, “London Calling? Long-Distance Connections in Australian Literature”, in Tully Barnett, Nena Bierbaum, Syd Harrex, Rick Hosking and Graham Tulloch, eds., *London Was Full of Rooms*, Adelaide: Lythrum Press (2006): 244-252, 245.

⁵² For a background on the AustLit: The Resource for Australian Literature, please visit <<http://www.austlit.edu.au:7777/presentations/index.html>>.

⁵³ AustLit, “Coverage”, online, <<http://www.austlit.edu.au/about/coverage>>, accessed 27 February 2009.

valid caution and it should signal the importance of being aware of the complexities and difficulties in any kind of empirical analysis that relies on datasets that, by their very nature, grow and change with the addition of new information. With reference to the Carter Brown Mystery Series and the Australian Common Reader projects, this appears an implicit understanding of all research drawing upon archival and bibliographic materials — even in those instances where the datasets appear to be complete. Therefore, as Joshi and also Katherine Bode claim, “rather than forcing a divide between ... statistics and cultural understanding, we should use one to enhance the other”.⁵⁴

The Australian Literary Disciplinary Space

In Australia, any project intending to apply computational power to the analysis and visualisation of book history data must eventually turn its attention to AustLit, the “Australian Literature Resource” (formerly “The Resource for Australian Literature”, 2006, and “The Australian Literature Gateway”, 2002).⁵⁵ As the largest holder of information correlated with Australia literature, AustLit represents a growing “structure of authority”⁵⁶ in the field of Australian creative and critical writing. Over time, it has established the cultural and institutional power to shape and set the legitimate definitions (and to influence the direction of bibliographic definition systems) for classifying Australian works and, more specifically, works as Australian. Collaborating with twelve Australian universities and the National Library of Australia, AustLit operates as a “networked digital research environment” building a web accessible “comprehensive bibliographic record of the nation’s literature”.⁵⁷

AustLit classifies works according to its own published scope policy, a process that might be described as the “imposition of a form of thought”⁵⁸ on a representative regime of works or

⁵⁴ Katherine Bode, “Beyond the Colonial Present: Quantitative Analysis, “Resourceful Reading” and Australian Literary Studies”, *Journal of the Association for the Study of Australian Literature*, online (2 May 2008): 189.

⁵⁵ *Internet Archive Wayback Machine*, online, <http://web.archive.org/web/*/http://www.austlit.edu.au/>, accessed April 2009.

⁵⁶ Pierre Bourdieu, *The Field of Cultural Production: Essays on Art and Literature*, Cambridge: Polity Press (1993).

⁵⁷ Kerry Kilner, “A Fool’s Errand? Or, Ten Thousand Authors and the Madness of a National Bibliography”, AustLit, online (2006), <<http://www.austlit.edu.au:7777/presentations/index.html>>, accessed April 2009.

⁵⁸ Jacques Ranciere, *The Politics of Aesthetics*, London: Continuum International Publishing Group (2006): 34.

as a process, which assesses texts against specific “frameworks of acceptance”.⁵⁹ AustLit’s primary aim is to “enhance and support research and learning in Australian literature”⁶⁰ and it achieves this through adapting online technologies to assist bibliographic discovery. So successful is this relationship between institutional power and the use of a web browser that AustLit’s bibliographic ascriptions appear on the computer screen as properties of the texts or works it has inspected. That is, at the level of on-screen interaction, the Australian Literature Resource operates as a database responsive to queries and as a system that requires its users to view search results as possessing considerable paratextual authority and rigorously authenticated details. Just as Jones has argued that “the [colonial library] catalogue ordered society’s body of knowledge within its card system”,⁶¹ today it might be argued that Australia’s literary knowledge is being shaped and organised by a website search form. In this way, as an internet-based resource, AustLit associates the power to say with authority what is an Australian work — a power traditionally held by human literary agents — with an interactive licensed technology product in return for an annual subscription calculated “under a range of pricing strategies”.⁶²

This distinguishes AustLit’s “canonical vision” as a “product of privilege”⁶³ within what Ken Gelder might call the “on-going canonisation of Australian literature through [a] well-funded, centralised editorial project”.⁶⁴ Certainly, literature from a researcher’s perspective has often been in a sense a product of privilege. One has only to price rare print bibliographies of Australian novels in first-hand and second-hand bookstores to recognise that contemporary print bibliographies continue this tradition.⁶⁵ However, if a humanities researcher is affiliated with a university or a public library that absorbs the subscription, it is reasonable to assume that they would be familiar with AustLit’s main website, or with what can be described as a virtual epistemic object constituted for academic consumption. This

⁵⁹ Edward W. Said, *Culture and Imperialism*, London: Chatto and Windus (1993): 376.

⁶⁰ Tessa Wooldridge, “Ensuring the Best - AustLit: The Resource for Australian Literature”, Conference Paper (Melbourne: Australian and New Zealand Society of Indexers, 20 March 2005).

⁶¹ Caroline V. Jones, “The Influence of Angus and Robertson on Colonial Knowledge” *Journal of the Royal Australian Historical Society* 89.1 (2003): 26-37, 31.

⁶² AustLit, “How to Subscribe”, online, <<http://www.austlit.edu.au/subscribe>>, accessed 27 February 2009.

⁶³ Pierre Bourdieu, *The Field of Cultural Production: Essays on Art and Literature*, Cambridge: Polity Press (1993): 19.

⁶⁴ Ken Gelder, “Recovering Australian Popular Fiction: Towards the End of Australian Literature” *Australian Literary Studies in the 21st Century: Proceedings of the 2000 ASAL Conference* (2001): 112-120, 118.

⁶⁵ In April 2009 all four volumes of the *Bibliography of Australian Literature* were on sale as a complete collection at the University of Queensland Press for \$499 AUD.

terminology is not intended to truncate AustLit's institutional and educational power, nor the oversight it exerts on the Australian literary disciplinary space. Indeed, it quickly becomes clear to any user of the AustLit website that its layered structured depth represents knowledge work produced by specialists and experts who routinely inspect and interpret bibliographic materials. Instead, as one of many key stakeholders in — and primary producers of — the contemporary Australian literary disciplinary space, the intent is to signal AustLit's contemporary cultural grammar; that is, its user-centred metaphor of "search". Furthermore, this is to link AustLit websites with broader postsocial⁶⁶ trends that aggregate information and expertise within a technological setting to produce a complex informational package, whose "objective properties"⁶⁷ are constituted under AustLit's brand,⁶⁸ available online as a discreet object of knowledge in its own right.

Case Study: *The Devil's Advocate* is in the Detail

The temptation in applying machine learning methods to humanities data is to interpret a computed result as some form of proof or determinate answer. In this case, the validity of the evidence lies inherent in the technology. This can be problematic when the methods are treated as a black box, a critic ex machina.⁶⁹

⁶⁶ The term postsocial reflects "the increased presence and relevance which non-human objects have assumed in contemporary life, and refer specifically to the kind of bonds which humans have developed with objects". Mayall, "Attached to their Style: Traders, Technical Analysis and Postsocial Relationships", p. 423. This is not to imply that scholars develop a relationship with an AustLit website though most Australian literary scholars will, at one point or another, turn to an AustLit website instead of a human expert. Rather, I wish to suggest more broadly that in the contemporary age scholars develop attachments to the virtual objects, online tools and internet-hosted information resources which help facilitate their particular brand of cognitive and interpretative labour. These "attachments" can take the mild form of bookmarking internet favourites to the not-so-mild form of hoarding massive amounts of data (or quantifiable mass) for the potential next new breakthrough that such a quantity of information must surely be concealing.

⁶⁷ C. Lury, "Marking Time with Nike: The Illusion of the Durable", *Public Culture* 11.3 (1999): 499-526, quoted in Mark Poster, "Consumption and Digital Commodities in the Everyday", *Cultural Studies* 18.2/3 (March / May 2004): 409-423, 421.

⁶⁸ AustLit's appearance and logo was created by Inkahoots, a graphic design company based in Brisbane which specializes in matching a client website's "look and feel" more strategically with its message, services and brand. <http://www.austlit.edu.au/about/acknowledgements>, accessed August 2010.

⁶⁹ D. Sculley and Bradley M. Pasanek, "Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities" *Literary and Linguistic Computing* 23. 4 (2008): 421.

Understandably and quite reasonably, AustLit has an investment in being able to advise how useful its informational resource will be to particular forms of analysis, especially those conducted outside the “singular symbolic surface”⁷⁰ of its websites, and the provenance of data available from AustLit to this end. In this regard, on sustained reflection, the caution of a literary resource being “insufficiently comprehensive ... for statistical analyses” should not be dismissed — at least not opportunistically.

This admission can be supplemented with two illustrations. Enacting St Clair’s argument in *The Reading Nation in the Romantic Period* that the exercise of enumerative bibliography might prove useful for literary and cultural history, at a mid-year 2007 conference I presented a sequence of images that applied statistical methods to AustLit tagged-text data. One table listed the top Australian novels reprinted internationally during the period 1890-2005. Drawing on the print-cultures logic that reprints can be a commercial indicator of demand, I presented this table alongside an argument that a picture, though oblique, could be built up of modern literary tastes and demands during the twentieth century through statistical analysis, specifically revealing which Australian novels publishers internationally reprinted or translated the most. The aim was not to solve or answer any particular problem about Australian publishing but rather — to recast Willard McCarty’s use of classicist Don Fowler as a redemptive personal motive — to make them worse, on the assumption that surprising and unusual results would create a context to ask new questions or refine existing ones.⁷¹

Over a two week period in June 2007 where I was fully occupied with writing computer instructions and mining the AustLit database, I engineered an algorithm or what McCarty refers to as the “black box” of “unexamined or obscure process[es]” underpinning any humanities-based computing project.⁷² As the algorithm behind the spread sheet of ranks, authors, years, works and totals, it grouped manifestations or reprints of a work with their

⁷⁰ Margery Mayall, “Attached to their Style: Traders, Technical Analysis and Postsocial Relationships” *Journal of Sociology* 43.4 (2008): 425.

⁷¹ Willard McCarty, “Humanities Computing”, *Encyclopaedia of Library and Information Science* (2003): 1224.

⁷² Willard McCarty, “Humanities Computing”, *Encyclopaedia of Library and Information Science* (2003): 1230.

primary (first edition) title.⁷³ This formed many subsets of the kind where, for example, *La mochila* (1956), *Un sacre petit paquet: roman Australien* (1957) and *Shirali: roman* (1978) are correlated with their central (often English-language) first published title, which in this example is *The Shiralee* (1955) by D'Arcy Niland. The algorithm chronologically ordered the manifestations within each subset, allowing also for the identification of first and last years of publication, and then counted the number making up each set. The results were collated in a tabular format and arranged from highest to lowest through the application of a ranking system keyed to subset totals, which in turn suggested a hierarchy of decreasing significance with the international translation of Australian novels (see Figure 1, page 325: all figures mentioned in this chapter can be found in Appendix A). In this way, rank one listed Colleen McCullough's 1977 work, *The Thorn Birds*, as having 47 manifestations during the years 1977 to 2005 and it could be reasonably interpreted to have more value (though what kind of value was not made clear) than, say, Brown's *The Unorthodox Corpse* which inhabited rank nine with only 18 manifestations internationally. This table was imported into that "ubiquitous form of digitally assisted demonstration",⁷⁴ PowerPoint, the Microsoft software product researchers and academics regularly employ to add persuasive power to their conference demonstrations. Within a larger narrative of seventeen images focusing on quantifiable outcomes, it was presented as slide number ten before a conference audience which shared scholarly interests in the disciplinary space of Australian literature.

These slides ranged from spread sheets representing Australia's most productive authors, to line graphs indicating publishing outputs throughout the twentieth century, to a final image of a NASA world map dotted with places of publication that signalled via a kind of GIS (geographic information system) where Australian novels have been produced throughout the planet (see Figure 6, page 327). Conjointly, this collection supported Stark's and Paravel's claim that PowerPoint enables the bringing together of "facts with different textures"⁷⁵ and its mention here is to invite awareness of the "technical and rhetorical modalities of digital demonstrations" which often prop up "the staging/screening of "facts",

⁷³ Jason Ensor, "Reprints, International Markets and Local Literary Taste: New Empiricism and Australian Literature", *Journal of the Association for the Study of Australian Literature* (May 2008): 208.

⁷⁴ David Stark and Verena Paravel, "PowerPoint in Public: Digital Technologies and the New Morphology of Demonstration" *Theory, Culture and Society* 25 (2008): 32.

⁷⁵ David Stark and Verena Paravel, "PowerPoint in Public: Digital Technologies and the New Morphology of Demonstration" *Theory, Culture and Society* 25 (2008): 44.

[and] their circulation”.⁷⁶ However, this use of PowerPoint was also an attempt to satisfy a (as then unacknowledged) personal drive to transport the conference audience to a “distant imaginary”⁷⁷ of diverse mathematical virtuosity and empirical certainty.⁷⁸ This was in the service of constituting “new genres for scholarship”⁷⁹ in Australian literature, of course, but it nevertheless blurred the analytic labour conventionally divided between scholars and technicians.⁸⁰ A version of slide ten was later published with the (necessary) apologetic that it was to be considered “provisional ... upon the completion of the AustLit database in the future and [antecedent to] the findings of a follow-up statistical analysis”.⁸¹

I did not think much more about slide number ten until February 2009 when another analysis of Australia’s top reprinted novels using the same algorithm was conducted (see Figure 2, page 325). Though some changes were expected, most surprisingly the only novels which remained familiar to both tables in the uppermost ranks were *Schindler’s Ark*, transitioning from rank 2 to rank 8, and *The Thorn Birds*, demoted from rank 1 to rank 5 by Morris West’s *The Devil’s Advocate* (with an impressive-looking 65 manifestations). If this study was still seeking a context to pose new questions, it had certainly found one in the lack of correspondence between these two tables. Such a lack would command any researcher to ask what happened between July 2007 and February 2009 to initiate such a dramatic reconfiguration of the publishing facts covering 115 years of Australian literary

⁷⁶ David Stark and Verena Paravel, “PowerPoint in Public: Digital Technologies and the New Morphology of Demonstration” *Theory, Culture and Society* 25 (2008): 34.

⁷⁷ David Stark and Verena Paravel, “PowerPoint in Public: Digital Technologies and the New Morphology of Demonstration” *Theory, Culture and Society* 25 (2008): 35.

⁷⁸ For data, as a component of methodology, “can scarcely be deployed without implicitly stating the affiliation of the speaker, and not as a mere matter of fact but as a declaration of kinship, vested interest, antagonism, defensiveness and so forth.” Julia Flanders, “Data and Wisdom: Electronic Editing and the Quantification of Knowledge” *Literary and Linguistic Computing* 24. 1 (2009): 54.

⁷⁹ Willard McCarty, “Humanities Computing”, *Encyclopaedia of Library and Information Science* (2003): 1230.

⁸⁰ On the traditional separation between scholars and technicians (since academics often employ RAs with the informational and database skill sets they themselves may not possess), Julia Flanders asks: “If the computer merely displays knowledge to a post-production society, what might this imply about our mechanisms for generating new (as opposed to retrieving and redeploying old) expert knowledge? How real is the danger that the scholar-worker, whose origins lie in a nineteenth-century conception of learning as heroic endeavour, will be transformed into the scholar-technician?”. Julia Flanders, “Data and Wisdom: Electronic Editing and the Quantification of Knowledge” *Literary and Linguistic Computing* 24. 1 (2009): 61.

⁸¹ Jason Ensor, “Reprints, International Markets and Local Literary Taste: New Empiricism and Australian Literature”, *Journal of the Association for the Study of Australian Literature* (May 2008): 201.

history and to ask, perhaps more significantly, what the methodological implications might be for future statistical analyses of this kind.

One immediate answer is that the foundational dataset upon which the enumerative bibliography was conducted had changed significantly with the addition of new information, enough to not only reorder the original 2007 findings but to significantly replace them with an altogether different list of works. (In fact, 4,202 manifestation records and 2,107 first edition records were added to the database for the period under analysis, 1890-2005. However, as an aside, to contextualise these numbers, according to the daily report that appears on the front page of its website, it is worth noting that AustLit reported 558,591 works in its database on 22 June 2007,⁸² compared to 626,376 works on 4 April 2009. This suggests that the whole database grew by approximately 67,785 works, of which the additional manifestation and first edition records under discussion represent only 10% or less of AustLit's total bibliographic growth between July 2007 and February 2009.)

This creates an ethical dilemma. On the one hand, it remains reasonable to stand by the claim that the empirical certainty reflected in each table is nonetheless accurate for the scope of data available *at the time of analysis*. From a data-mining perspective, the statistical results were calculated in a valid manner. On the other hand, the table for July 2007 (which reflects the processing of over 14,750 manifestations, a not insignificant amount) is correct — and yet now obviously incorrect — and the table generated in February 2009 (representing the processing of 18,954 manifestations) is also correct *as of writing*. It too, however, will eventually cycle through its “half-life” of certainty much like its 2007 predecessor.

Borrowing a term from the glossary of nuclear physics, by “half-life” its definition refers to the time in which half the conclusions of a particular set of academic findings disintegrate. This half-life is a characteristic property of all research including the hard sciences, which are often popularly considered incontrovertible. However, this half-life is a particularly important caveat in enumerative bibliography. Here, the perception of change in knowledge may be measured in briefer time-frames than, say, traditional scholarship which relies on archival documents (whose retrieval and synthesis into new historical facts requires a period of activity considerably longer than the time it takes to unleash an

⁸² *Internet Archive Wayback Machine*, online,
<http://web.archive.org/web/*/http://www.austlit.edu.au/>, accessed April 2009.

algorithm on updated datasets). The underlying difficulty is that most scholars function within university environments keen for quantifiable research contributions. It is a generally accepted working condition that academics publish their findings as soon as practically possible and therefore typically it is an unsatisfactory situation to refrain from being issued in peer-reviewed publications. In light of the above paradox — where the half-life of new empiricist analysis is likely in some modes to be less than the time which passes between the acceptance of an article by an editor of a journal and its eventual publication — is presenting a study's conclusions as subject to qualification and on-going work (again, those necessary apologetics) sufficient insurance against the risk of one day being out of date but not out of print?

It is appropriate at this point to refer back to Jessop who writes that the incomplete record is “a significant weakness of digital visualisation which will have to be addressed if its scholarly status is to be ensured”. Moreover:

Visual sources present the viewer with a complete, and convincing, picture that is often derived from an incomplete record but the nature of the media used requires that the gaps be filled during its creation and thus concealed. If the applications of representation and abstract secondary sources are to be regarded as anything more than mere entertainment it must be ensured that viewers are aware of not only what is present but also what is omitted and the levels of uncertainty of that which is present.⁸³

The case study above illustrates the ease with which this kind of problem can be encountered in a quantified analysis of Australian publication data. Whilst the availability of PowerPoint, statistical analysis packages, internet-hosted empirical tools, and online database resources confer a level of scientific authenticity to humanities knowledge production and outcomes, the foreshortened half-life of some computer-based research

⁸³ Martyn Jessop, “Digital Visualisation as a Scholarly Activity” *Literary and Linguistic Computing* 23.3 (2008): 287.

should encourage caution and an overarching method to contextualise findings. One way to guard against such gaps is to create a set of standards reflecting good practice under an *Australian Charter for the Computer-Based Representation of Literary History*, within whose context new findings would be presented. This would be achieved through an open debate using *The London Charter* as a template. It would: “provide a benchmark having widespread recognition among stakeholders, promote intellectual and technical rigour ..., ensure that computer-based visualisation processes and outcomes can be properly understood and evaluated by users, [and] enable computer-based visualisation authoritatively to contribute”⁸⁴ to the study of Australian literary history and Australian print cultures. It would also respond to what Sculley and Pasanek identify as a “need to find an articulate consensus on meaningful standards for experimental evidence provided by data mining”.⁸⁵ Additionally it is recommended, should an *Australian Charter* (or *London Charter Implementation Guidelines for Australian Literature*) eventuate, that it would respond to the following four issues that are specific to Australian literature:

Issue One: Black Boxed Analysis

There are two “black boxes” embedded within any humanities computing project analysing publication data. One is the code employed to process the analysis, the other is the dataset used in the analysis, both of which can be difficult to release to the public or a shared disciplinary / methodological commons but which need to be more open in order to be tested, challenged and incorporated by alternative, even competing projects. As Beckmann notes, part of the success of the hard sciences is that they “subsidize opposing voices”.⁸⁶ Thus, if technical questions are entangled with political questions over data ownership and access then it is perhaps beneficial to address both kinds of questions in parallel in order to advance methods of (and to encourage a healthy ecology of) quantitative analysis for Australian literature. That means debating sensitive issues of ownership, independent testing, reproducible methods and gate-keeping practices regarding data retention and knowledge creation, at least within the context of creating a shared disciplinary /

⁸⁴ “Objectives”, *The London Charter for the Computer-Based Visualisation of Cultural Heritage* (Draft 2.1), online (2009), <<http://www.londoncharter.org/>>, accessed April 2009.

⁸⁵ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 421.

⁸⁶ Roger Beckmann, “Literature, Science and Economic Rationalism”, in Katherine Barnes and Jan Lloyd-Jones, eds., *Words for their Own Sake: The Pursuit of Literature in an Economic Rationalist World*, Melbourne: Australian Scholarly Publishing (2004): 69.

methodological commons or online archive.⁸⁷ It follows, as McCarty suggests, that humanities computing “challenges issues of ownership, which is to say, reveals that many [source materials] are held in common and there is much to be gained from sharing them. If its real potential is understood, humanities computing can be quite threatening to the status quo”.⁸⁸

Issue Two: Data and Cultural Work

The creation of a data source is cultural work and especially so with data that is centred on culture or nation as its core organising concept. Although Australian literature is a rather welcoming environment for works and authors from around the world (mirroring on the one hand the “sign of the postcolonial”⁸⁹ and on the other hand, the de-centred and de-territorialising logics of capitalism), there is no escaping the key disciplinary conceit that every entry in the AustLit database is taken to be importantly correlated with Australia. Yet in assessing a book’s suitability for inclusion, “we “don’t just peer” ... [w]e must also “interfere” with the incoming data based on what we know we are trying to observe”.⁹⁰ For example, transgressing borders seems an inevitable issue for any data-organising principle used in Australian literature. In a survey of Australian everyday cultures, Bennett, Emmison and Frow acknowledge the “difficulties in focussing on the ... origin” of authors when they sought insight into the reading tastes of their respondents.⁹¹ For them, many writers “are truly international in the sense that they reside in more than one country at different times of the year, or they may have moved permanently from their country of origin to reside elsewhere”.⁹² We might recognize this as the “paradox of authenticity in the age of postmodern travel” and multinational companies.⁹³ This is also a view that Macmillan, the

⁸⁷ See point “6.3.4 Make data available and methods reproducible”, D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 423.

⁸⁸ Willard McCarty quoted in Stuart Moulthrop, “Computing, Humanism, and the Coming Age of Print”, *Seminar: Is Humanities Computing an Academic Discipline?*, online (3 December 1999), <<http://iat.ubalt.edu/moulthrop/essays/uva99/>>, accessed April 2009.

⁸⁹ Graham Huggan, *Australian Literature: Postcolonialism, Racism, Transnationalism*, Oxford: Oxford University Press (2007): 150.

⁹⁰ Willard McCarty, “Digitising is Questioning, or Else”, Long Room Hub Lecture Series, Trinity College Dublin, online (16 April 2008), <<http://staff.cch.kcl.ac.uk/~wmccarty/>>, accessed April 2009.

⁹¹ Tony Bennett, Michael Emmison, and John Frow, *Accounting for Tastes: Australian Everyday Cultures*, Cambridge: Cambridge University Press (1999): 213.

⁹² Tony Bennett, Michael Emmison, and John Frow, *Accounting for Tastes: Australian Everyday Cultures*, Cambridge: Cambridge University Press (1999): 213.

⁹³ Inez Baranay, “Multiculturalism, Globalisation and Worldliness: Origin and Destination of the Text”, *JASAL* 3 (2004): 129.

publisher of Alan Yates' autobiography *Ready When You Are*, CB would agree with.⁹⁴ Yates, as the author behind the extraordinarily successful pulp literature alias "Carter Brown", published extensively in Australia by the New South Wales firm of Horwitz, is described as "Australia's own and America's own and Britain's own"⁹⁵ — any attempt to confine him as Australian only is a "vigorous assertion".

Yate's designation as an Australian writer is of particular significance in any statistical approaches to Australian literature. London-born, Yates arrived in Australia at the age of twenty-three, after which he wrote detective fiction for nearly two decades before returning to England in 1967. Yates was still living in London when his autobiography was published in 1983 but AustLit records his death just two years later in New South Wales. It is clear, to interpret his movement back to Australia and to quote from his autobiography, that Alan Yates retained a "great deal of affection for Australians".⁹⁶ This statement by Yates is important because even in his fifties he remained fond of Australia but did not explicitly identify himself as Australian. Nonetheless, because Yates is co-opted as an Australian writer, as will be observed in chapter three his impact on any statistical analysis of Australian literature is substantial and accounts for a considerable percentage of New South Wales publication output in the 1960s. If Yates' status was to change, so too would any empirical view of Australian history. Though Macmillan's dust-jacket comments are an instance of publishers amplifying a writer's significance, in the market-hyped sense that the author proverbially "belongs to the world" rather than any limited group of people or single place, their statements and Yates' own draw attention to the problems of thinking about authors, books and their relationships to groups of readers and individual places.

Yates and his Carter Brown alias do not figure in Bennett, Emmison and Frow's study of taste but the issue of linking authors to specific groups and places continues. What becomes important then to Bennett *et al's* research is "the content of reading material rather than the nationality of authors".⁹⁷ Though having reservations, Bennett *et al* claim it

⁹⁴ Alan Yates, *Carter Brown, Ready When You Are, C.B.! The Autobiography of Alan Yates Alias Carter Brown*, Melbourne: Macmillan (1983).

⁹⁵ Alan Yates, *Carter Brown, Ready When You Are, C.B.! The Autobiography of Alan Yates Alias Carter Brown*, Melbourne: Macmillan (1983).

⁹⁶ Alan Yates, *Carter Brown, Ready When You Are, C.B.! The Autobiography of Alan Yates Alias Carter Brown*, Melbourne: Macmillan (1983): 1.

⁹⁷ Tony Bennett, Michael Emmison, and John Frow, *Accounting for Tastes: Australian Everyday Cultures*, Cambridge: Cambridge University Press (1999): 212.

is “necessity” which pushes them to “pragmatically assign”⁹⁸ the country-of-origin categorisation for some authors. Reference is made, for example, to Peter Carey who is coded as Australian, though it is acknowledged he has been living in the United States for some time, and through this the issue seems closed. However, if two titles by recognised British author D. H. Lawrence — *Kangaroo* (1923) and *The Boy in the Bush* (1924) — are assessed by AustLit to be Australian novels because of their “reading content” or setting, then the issue is actually further problematised by the *Accounting for Taste* example.

It would seem books and authors, as Andrew Hassam argues with regards to a writer’s national identity, “can be regarded as Australian despite one’s citizenship, place of birth or where one lives: the important factor is one’s association with Australia”.⁹⁹ Thus, while close readings of an AustLit record might unpack a work’s sometimes invisible or even obscure connection to Australian literature, a distant reading of the database does not have this cognitive power and therefore cannot account for the bibliographer’s judgement behind the creation of the data entry. This can distort some historiographic conclusions derived from quantitative analysis. That is, an Australian literary database, like Ramsay claims for software, “cannot be neutral ... since there is no level at which assumption disappears” nor where a “demonstrably non-neutral act of interpretation can occur”.¹⁰⁰ Ways then need to be discovered to reveal this stored labour of bibliographic assignment (which endows Australian literature with much of its power). As Stuart Moulthrop remarks: “Data is past participle, that which is given, but in the humanities we tend not to accept the given without scepticism or inquiry”.¹⁰¹ Indeed, should we take such things at their word and be done with critical inquiry?¹⁰² Sculley and Pasanek conclude in their study of data mining in the humanities that “we must pay strict attention to the manner in which the data sets are constructed”.¹⁰³ By identifying and evaluating research sources in a “structured and documented way”, future studies incorporating new empiricism may explore how Australian

⁹⁸ Tony Bennett, Michael Emmison, and John Frow, *Accounting for Tastes: Australian Everyday Cultures*, Cambridge: Cambridge University Press (1999): 213.

⁹⁹ Andrew Hassam, “From Ned Kelly to Don Bradman: India, Australia and the Incongruities of Globalisation”, in C. Vijayasree, R. Azhagarasan, Bruce Bennett, Mohan Ramanan, R. Palanivel and T. Sriraman, eds., *New Bearings in English Studies: A Festschrift for C.T. Indra*, Hyderabad: Orient Longman (2008).

¹⁰⁰ S. Ramsay, “In Praise of Pattern”, *Text Technology* 2 (2005): 182.

¹⁰¹ Stuart Moulthrop, “Computing, Humanism, and the Coming Age of Print”, *Seminar: Is Humanities Computing an Academic Discipline?*, online (3 December 1999), <<http://iat.ubalt.edu/moulthrop/essays/uva99/>>, accessed April 2009.

¹⁰² A question posed by a reader of an earlier version of this essay.

¹⁰³ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 412.

literary data and “visual sources may be affected by ideological, historical, social ... and aesthetic” factors.¹⁰⁴

Issue Three: Critical Awareness and Stance

As with the technologies of print, which took centuries to stabilise into the forms exploited today, we must be highly conscious of the “newness” of new empiricism because, as is also claimed about PowerPoint and which is applied to computing in Australian literature, we are using “technology in its early moments of adoption, during which there are important questions about when and, if so, how it becomes stabilised”.¹⁰⁵ While it is useful to link contemporary humanities computing to previous projects of exemplary empirical studies that pre-date information technology, this can have the effect of naturalising computational variants of empirical research before they have fully unfolded within humanities disciplines, leaving critical awareness of the influence of technology — of its orchestrated effects — on methodologies a little weak.

Issue Four: The Half-Life of Research

As discussed above, conclusions drawn via the computational turn can be prone to having a much shorter half-life (that is, the lapse of time before new findings are presented which challenge and overturn previous knowledge claims) than traditional scholarship preoccupied with archival materials. Documenting knowledge claims would make clear to other scholars what a particular visualisation of quantified analysis is seeking to represent, and the extent of any factual uncertainty. Connected with this point, different levels of scale in distant reading have proportional effects on claims of certainty when the foundational dataset is still maturing. A form of distant reading that interprets trends (as in Figures 3 and 4, page 326) can be reasonably expected to outlive a microcosmic form that parses “top ten”-type enquiries. Figures 3 and 4, though also generated from the same AustLit snapshots taken nineteen months apart as the top Australian reprint lists under discussion, show very little observable difference in their representation of Australian novel production — certainly nothing that overturns the July 2007 reading of Australian versus English publication trends — and therefore are findings still in print that have not yet disintegrated. (But, it should be

¹⁰⁴ “Research Sources”, , *The London Charter for the Computer-Based Visualisation of Cultural Heritage* (Draft 2.1), online (2009), <<http://www.londoncharter.org/>>, accessed April 2009.

¹⁰⁵ “Research Sources”, , *The London Charter for the Computer-Based Visualisation of Cultural Heritage* (Draft 2.1), online (2009), <<http://www.londoncharter.org/>>, accessed April 2009.

noted, there is still a significant difference of 1,188 first edition titles between the totals of figures 3 and 4 which would impact other kinds of analysis comparing Australia with Britain.) Microcosmic enquiries like “top ten” lists, though intriguing, can be a kind of secondary instrumentalisation which is brought about when a dataset’s *stabilised-for-now* status is taken to be broadly indicative of a future unchanging or *stabilised-enough* nature.¹⁰⁶ As my experience testifies, greater detail does not automatically correlate with accuracy.¹⁰⁷ Not accounting for secondary instrumentalisation invites the possibility of a foreshortened half-life in the presentation of any microcosmic research outcomes. Additionally, other kinds of microcosmic enquiries and reports — like “how many times does etc?” — run the risk of being received as quick journalistic facts with short-lived historical resonance, even though they may remain what Priya Joshi says of all statistics: “lies that tell a truth that would not otherwise be evident”.¹⁰⁸ Such lists and reports, if orphaned from critical thinking, should be revealed as a naive form of new empiricism. Indeed Sculley and Pasanek caution professional readers like literary scholars that “just because results are statistically valid and humanly interpretable does not guarantee that they are meaningful. ... [For] we can give a gloss or a paraphrase for all varieties of nonsense”.¹⁰⁹ The core issue then is to recognise that “some representations are better than others ... in the sense of providing a more useful analytical model”¹¹⁰ for cases where the dataset may not be relatively stable.

¹⁰⁶ I would argue that “top-ten”-like enquiries are generally susceptible to this kind of problem when working with datasets that have not stabilised. For example, with the 2007 dataset used in my ASAL presentation, I found that the top ten publishers of first edition Australian novels worldwide for the period 1960-2005 were: (1) Cleveland, (2) Horwitz, (3) Mills and Boon, (4) Penguin, (5) Allen and Unwin, (6) Angus and Robertson, (7) Scripts, (8) University of Queensland Press, (9) Robert Hale and (10) Pan Macmillan Australia. In 2009, the same analysis applied to the most recent data snapshot yielded: (1) Cleveland, (2) Horwitz, (3) Mills and Boon, (4) Penguin, (5) Scripts, (6) Robert Hale, (7) Angus and Robertson, (8) Allen and Unwin, (9) University of Queensland Press and (10) Pan Macmillan Australia. Though minor, in comparing the analysis applied to data snapshots nineteen months apart, there were nonetheless shuffles in order for ranks 5, 6, 7, 8 and 9. While this does not negate the overall illustrative and indicative value of this kind of enquiry, it does confirm that caution should be exercised in its presentation within an argument.

¹⁰⁷ Julia Flanders, “Data and Wisdom: Electronic Editing and the Quantification of Knowledge” *Literary and Linguistic Computing* 24. 1 (2009): 56.

¹⁰⁸ Priya Joshi, “Quantitative Methods, Literary History”, *Book History* 5 (2002): 264.

¹⁰⁹ D. Sculley and Bradley M. Pasanek, “Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities” *Literary and Linguistic Computing* 23. 4 (2008): 420.

¹¹⁰ Julia Flanders, “Data and Wisdom: Electronic Editing and the Quantification of Knowledge” *Literary and Linguistic Computing* 24. 1 (2009): 56.

Conclusion: What then is the Worth of a Picture?

As arguably one of the first studies to extend new empiricism into an Australian context, applying computational methods like Moretti's to interrogate AustLit and construct a history of the publication of Australian novels over a hundred years, this chapter has discussed the shortcomings of some of the data. It has raised methodological concerns regarding the processing of bibliographic data and the presentation of quantitative analysis in literary-historical and print-cultural contexts. In the discussion of graphing and data visualisation, it agrees with Jessop's argument that "every representation ... is an effort to structure an argument and as such it is a rhetorical device".¹¹¹ This chapter positions AustLit as one of the most comprehensive and authoritative sources on Australian novels while noting the limitations, difficulties and complexities in working with publication data under critical-interpretative contexts. It recommends a future adaption of the *London Charter* as a framework for sustaining intellectual rigour in Australian applications of new empiricism in the humanities.

That information correlated with Australia's literary estate has continued to be added, preserved or discarded points to the dynamic rather than static nature of bibliographic practice. In Australian literature, a history of these bibliographic changes would constitute a window into the discipline's evolving relationship to Australian notions of identity, culture, and nation. In Australian print cultures, a history of these bibliographic changes would point to the international and transnational nature of textual production. It is these twin issues — of categorisation as cultural work and change over time within foundational datasets — which invite caution in statistical measurement even as each issue reveals that the creation, production and manufacture of text "functions *through* national boundaries".¹¹² On this view, the present study sustains David Carter's concept of the nation as "a political and cultural formation around which value and meaning are accrued"¹¹³ but within the critical recognition that a history of Australian publishing or Australian literature must take account of a "shifting set of relationships — between local, regional, national and international co-

¹¹¹ Martyn Jessop, "Digital Visualisation as a Scholarly Activity" *Literary and Linguistic Computing* 23.3 (2008): 287.

¹¹² David Carter, "Good Readers and Good Citizens: Literature, Media and the Nation", *Australian Literary Studies* 19 (1999): 149.

¹¹³ David Carter, "Good Readers and Good Citizens: Literature, Media and the Nation", *Australian Literary Studies* 19 (1999): 138.

ordinates".¹¹⁴ The next chapter will demonstrate that the large-scale quantitative analysis of AustLit data usefully tests disciplinary boundaries¹¹⁵ and creates representations through which we might gain new understandings of the past.¹¹⁶ Such analysis can provide a more detailed overview of the national and international coordinates within the production of Australia's literary estate.

In conclusion, AustLit is a complex and constructed object that presents knowledge of Australia's literary past in digital form.¹¹⁷ It is the leading edge of an ongoing project whose work stretches back to foundational publications by J. A. Ferguson, E. Morris Miller and H. M. Green.¹¹⁸ However, AustLit is also a second-order representation of the developing postsocial relationship between humanities scholars and the use of virtual objects in research. In *Principles of Literary Criticism*, I. A. Richards, credited with pioneering the literary movement of "Practical Criticism", began his seminal work with "A book is a machine to think with".¹¹⁹ Today, eighty-five years since the first publication of *Principles of Literary Criticism*, we increasingly use machines to think about books but, to contemporise the second half of Richards' opening formulation, these need not usurp the humanities scholar. In view then of the above points and with reference to Figure 5 (page 327), this chapter appears to insist on a lot of ground-work before answering its title question, "is a picture worth 10,175 Australian novels?". Perhaps an answer is finally possible but within the context of an *Australian Charter for the Computer-Based Representation of Literary History*. Through this, though *Places of Publication* would be unable to make any claims to truth or historical fact, it would nonetheless be able speak its answer with a greater level of

¹¹⁴ David Carter, "Good Readers and Good Citizens: Literature, Media and the Nation", *Australian Literary Studies* 19 (1999): 145.

¹¹⁵ See, for example, "Australian literary studies: boundary work and the new empiricism", *bildungs & food*, online (January 2009), <<http://eurhythmaniac.wordpress.com/2009/01/24/australian-literary-studies-boundary-work-and-the-new-empiricism/>>, accessed April 2009.

¹¹⁶ Flanders makes this point with regards to the analytic modeling of a text's structure through multiple iterations. The point however is equally valid for data analysis. Julia Flanders, "Data and Wisdom: Electronic Editing and the Quantification of Knowledge" *Literary and Linguistic Computing* 24. 1 (2009): 62.

¹¹⁷ To rewrite Hayot and Wesp's use of Edward Said in their discussion of virtual worlds. Eric Hayot and Eric Wesp, "Towards a Critical Aesthetic of Virtual-World Geographies", *Game Studies* 9.1 (April 2009): 1. Online, http://gamestudies.org/0901/articles/hayot_wesp_space, accessed May 2009.

¹¹⁸ J. A. Ferguson, *Bibliography of Australia*, Sydney: Angus & Robertson (1941-69); E Morris Miller, *Bibliography of Australian Literature 1795—1938*, Melbourne: Melbourne University Press (1940); H M Green, *A History of Australian Literature*, Sydney: Angus & Robertson (1961).

¹¹⁹ I. A. Richards, *Principles of Literary Criticism*, London and New York: Routledge Classics (2001): vii.

confidence in the face of datasets being “insufficiently comprehensive”, as a better-founded hypothesis, with some probabilities of certainty.