A growing number of data-rich analyses of literature and literary culture—variously described as ‘distant reading’ (Moretti 2005), ‘algorithmic criticism’ (Ramsay 2008), ‘macroanalysis’ (Jockers 2013), and ‘new empiricism’ (Bode and Dixon 2009)—have in the last decade significantly transformed literary studies. This international trend is strongly reflected in Australian literary studies, where there have been multiple quantitative analyses of borrowing records (Dolin 2004; 2006; Lamond and Reid 2009), book sales (Davis 2007; Zwar 2012a; 2012b), newspaper reviews (Thomson and Dale 2009), and bibliographic data (Bode 2010; 2012a; 2012b; Carter 2007; Ensor 2008; 2009; Nile and Ensor 2009). As important as this work has been for reconceptualising the object and scope of literary studies, its credibility and progress as a whole is inhibited by the fact that many of these authors provide little detailed discussion of the processes involved in creating, curating, and analysing their data sources. Even less rarely do they publish these sources. While there are exceptions,¹ such lack of access to data is true of the most high-profile work in this field—including Jockers’ and Moretti’s influential monographs—and prevents other scholars from investigating, extending, and potentially challenging these authors’ findings and arguments.

This situation cannot be blamed on traditional publishing conventions. Not only are there now humanities journals that enable and encourage this activity—not least of all, digital humanities journals such as Literary and Linguistic Computing and Digital Humanities Quarterly—but any researcher capable of producing and analysing data in the first place can easily release this data online, and refer or link to it in any publication, print or digital.² The situation instead arises from the complex relationship of literary studies and methodology. Perhaps more than any other discipline, literary critics routinely occlude methodology, such that a ‘close reading’ of, for instance, sexuality in the works of Shakespeare and Marlowe is presented as a clear and apparent object of study without reference to the selections, and gradual accretion of evidence and patterns, that led to the identification and conceptualisation of this topic. More pointedly, those areas of literary studies—such as bibliography and scholarly editing—that focus explicitly on methodology have, in the last half century, been separated from and regarded as ancillary to the more high-profile and esteemed work of literary criticism and theory (McGann 2004, 409).

Arguably, the history of computational literary studies compounds this perception of methodologically explicit literary research as marginal and inferior, further diminishing the willingness of literary scholars in digital humanities to discuss and—by publishing their data—demonstrate their methodology. As Stephen Ramsay argues, part of the reason why traditional forms of humanities computing, such as computational stylistics, have been ignored by mainstream literary scholars is that their focus on method is bound up in a narrow, positivist scientific paradigm where ‘the accumulation of verified, falsifiable facts’ is the point of the exercise and the only ‘basis for interpretive judgment’ (2008). This approach is not only theoretically naïve but, existing at odds with the hermeneutics of most literary criticism, does not
address and advance the concerns of that field. In the shift from humanities computing to digital humanities, and the forging of new historical, interpretive, and expansive approaches to literary data, there is a sense in which the detailed attention to methodology that characterises computational stylistics has been sloughed off as part of an embarrassing history of scientific positivism.

While methodological openness and access to data are necessary for digital and data-rich literary scholarship to progress and gain credibility, it would be entirely counterproductive if this was the only focus; the field also needs to fulfil the requirements of all humanities research: of presenting arguments, providing insights, and advancing new understandings. This paper presents one model for integrating these aspects of research that draws explicitly on our interdisciplinary collaboration between a literary scholar and an astrophysicist to adapt for digital humanities a common form of publishing in astronomy. Key to research in this scientific field are articles that announce the release of new data; describe the processes, decisions, and selections underpinning the creation and exploration of this data; and present initial analyses and findings. Other astronomers are able to assess how the data has been collected, defined, and analysed: if they agree with this process and its outcomes, they can use the data in their own research, citing the original paper so that they do not need to go into the same level of detail about the data; if they disagree, they can challenge the findings and offer alternative interpretations. Importantly, such articles also provide a means of recognising and rewarding the important and time-consuming work of gathering, processing, and investigating large amounts of data, and in this respect could help to address a difficult problem in the digital humanities—the lack of reward for creating, implementing, and maintaining digital collections and resources— that has itself contributed to minimal publication of data sources in this field.

Following this model, in this chapter we announce a new resource for humanities research: five datasets regarding critical attention to Australian novelists. The first part of this chapter describes the processes by which we gathered data from AustLit—an online bibliography of works on and about Australian literature—and checked, cleaned, and analysed these datasets to identify their limitations and potential. The second part uses these datasets to explore canon formation in this national literary field, as well as related issues of cultural value of interest to the humanities broadly. Far from reasserting the narrow scientific positivism that Ramsay criticises in computational stylistics, the view of data in this chapter—and indeed, despite the perceptions of some humanities scholars, across the sciences—is as an outcome of multiple decisions, and based on arguments, not certainties, regarding degrees of reliability, bias, and purpose. Our point in this chapter is not only to combine the presentation of important research findings and questions with explicit discussion of methodology and publication of datasets; it is to highlight that these supposedly technical issues are themselves inherently critical and theoretical, and they cannot be separated off from the arguments they underpin. By describing our journey from the original conceptualisation of the project to the types of questioning we ultimately believe these datasets can support, we demonstrate that data-rich research requires, not comprehensiveness or completeness, but a clear understanding of data characteristics and parameters.

Methods

Created in 1999, AustLit merged a number of existing specialist datasets and bibliographies and has since involved well over a hundred individual researchers, supplementing and updating records of works of and about Australian literature. The database now contains almost a million records and aims to be the definitive virtual research environment and information source for
Australian literary, print, and narrative culture scholars, students, and the public. Indeed, due to the work of many individuals, the longstanding investment of government and institutional funding, and the relatively recent origins of Australian literature, AustLit is the most comprehensive online bibliography of a national literature currently available. In this respect, it offers a valuable test case for future quantitative work with a range of digital bibliographies not currently at this stage of development. At the same time, the data in AustLit is not perfect or complete; AustLit also warns that its data is not sufficient for statistical analysis. We return to these two points in detail.

Our collaboration began as an attempt to explore ‘The Critical Careers of Australian Novelists’, with respect to publication, reception, and reputation. AustLit lists ‘works by’ and ‘works about’ each Australian author in the database. Approximately half of all Australian authors with novels published from 1900 to 2006 (the period we consider) have no ‘works about’. Even so, ‘works about’ are by far the most extensive component of the database. Among these, articles in newspapers, literary magazines and academic journals, and books and book chapters predominate, but other forms include bibliographies and collections, correspondence, dissertations, literary responses and screen adaptations, and multimedia and websites. When we started the project, we were interested in whether different relationships between novel publication and critical reception exist for particular groups of novelists: for instance, do authors who receive no critical attention tend to write in particular genres, or at particular times; do men and women novelists tend to accrue critical attention at different rates; is there a particular trajectory of critical reception (for instance, immediate attention or the gradual accrual of critical acclaim) more likely to lead to an established and longstanding prominence in the critical arena (in other words, how might canonical literary reputations develop)? We hoped that analysis of large numbers of reviews of authors’ works would produce patterns representative of particular relations to the literary field and its processes of reputation building.

AustLit is designed, and predominantly used, for finding information about particular authors or texts. Its guided search function enables searching of particular literary forms, date ranges, genres, and so on, and up to 1,000 results can be displayed and extracted as tagged text; but there is presently no facility to explore the database using a general query language. The impossibility of directly querying AustLit, and the volume of data on critical works, meant that to perform this study we had to automatically extract the data by writing code that performed, much more quickly than is manually possible, the searches supported by AustLit (entering each author’s name, requesting ‘works about’ them, then going to each of those works to gather the bibliographical data). This process occurred in February 2007, using a list of novelists that excluded ‘non-AustLit’ authors, and yielded a dataset with characteristics shown in Table 1.

Table 1. Properties of the data set extracted from AustLit.

<table>
<thead>
<tr>
<th>Property</th>
<th>Count</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Number of Authors</td>
<td>11,305</td>
<td></td>
</tr>
<tr>
<td>Authors with $\geq$ 1 novels</td>
<td>5,837</td>
<td></td>
</tr>
<tr>
<td>Authors with $\geq$ 1 poems</td>
<td>5,463</td>
<td></td>
</tr>
<tr>
<td>Authors with $\geq$ 1 short stories</td>
<td>1,527</td>
<td></td>
</tr>
<tr>
<td>Total Number of Novels</td>
<td>19,671</td>
<td></td>
</tr>
<tr>
<td>Total Number of ‘Works About’</td>
<td>92,250</td>
<td></td>
</tr>
<tr>
<td>Review</td>
<td>47,067</td>
<td>(51%)</td>
</tr>
<tr>
<td>Criticism</td>
<td>18,164</td>
<td>(20%)</td>
</tr>
<tr>
<td>Biography</td>
<td>9,186</td>
<td>(10%)</td>
</tr>
<tr>
<td>Category</td>
<td>Count</td>
<td>Percentage</td>
</tr>
<tr>
<td>------------------</td>
<td>-------</td>
<td>------------</td>
</tr>
<tr>
<td>Column</td>
<td>7,630</td>
<td>(8%)</td>
</tr>
<tr>
<td>Interview</td>
<td>2,088</td>
<td>(2%)</td>
</tr>
<tr>
<td>Uncategorised</td>
<td>1,820</td>
<td>(2%)</td>
</tr>
<tr>
<td>Autobiography</td>
<td>1,492</td>
<td>(2%)</td>
</tr>
<tr>
<td>Correspondence</td>
<td>1,256</td>
<td>(1%)</td>
</tr>
<tr>
<td>Other</td>
<td>3,547</td>
<td>(4%)</td>
</tr>
</tbody>
</table>

As with any manually created resource, some entries in AustLit contain inconsistencies—such as the same information entered in different ways or under different headings—that present difficulties for automatic extraction. To check that our extraction process did not introduce errors, we conducted a number of tests, including randomly selecting subsamples of authors and works and manually comparing their attributes with AustLit’s online records and producing diagnostic plots to map the dataset to identify outliers. Outliers often reveal errors in the data, but if not erroneous, they can indicate exceptional or interesting cases. An example diagnostic plot is Figure 10.1, which displays authors according to the number of novels written versus the number of ‘works about’ received. By repeating these diagnostic procedures, we improved our system to the point where we were confident that the extraction process had not distorted the dataset.

Figure 10.1. Australian Novelists—Number of Novels Written and Number of ‘Works About’

From this point, we turned to assessing the nature and reliability of our extracted data. As might be expected with a database designed for individual searches, AustLit’s descriptions of its data are not helpful in determining the types of quantitative analysis it can support. Describing the ongoing work ‘to correct unevenness and gaps in bibliographic coverage’, AustLit states that ‘until this process is complete, some AustLit data is insufficiently comprehensive to be used for statistical analyses’. Even where the database attempts to provide detail, information about the
scope and content of AustLit’s ‘works about’ Australian authors and texts is frustratingly vague. Noting that its ‘coverage is stronger in some areas than in others’, AustLit continues, ‘Periodical and newspaper literature, for example, is much more comprehensively covered from 1988 than for earlier years. Some specialist areas are rich and virtually complete, while others continue to evolve. Coverage of the nation’s major newspapers is extensive but not exhaustive’. While this reference to 1988 proved useful in our analysis, as we will show, for the most part these descriptions—no timeframe for completion of the process of correction; vague references to ‘some’ and ‘other’ areas—offer no useful overview of content and scope for defining the parameters for quantitative analyses.

While, on the surface, AustLit is simply warning researchers of possible gaps in its records, at a deeper level, its reference to a future where these problems will no longer exist implies the possibility of perfect and complete data, and identifies such perfection and completeness as a precondition of quantitative analysis. The simple practical consideration of the inevitable lag between the ongoing publication of works of and about Australian literature and their inclusion in AustLit precludes this imagined future ever arriving. More importantly, ontologically, it is impossible for any dataset of Australian literature to be complete and correct. As is acknowledged elsewhere in AustLit, ‘The definition of “Australian” and “Literature” moves according to current debates and changing reading, teaching and research patterns’. Instead of a situation where AustLit approaches and eventually attains a reflection of the ‘real world’ of Australian literature, the database is engaged in an ongoing process of representing and constructing that category. Where AustLit can develop a set of parameters to define, for instance, an ‘Australian author’—including, at present, such considerations as where they were born, where they spent their formative years, and the content of their fiction—this category is inevitably dependent on a current set of interpretations and thus will change as the Australian literary field is continually redefined.

Rather than awaiting perfect data, our aim and challenge were to define a ‘static’ dataset suitable for analysis from a changing and evolving database; in this respect, characterising the data is key. More broadly, such characterisation is increasingly important for humanities research because of the growing interest in data-rich analysis of major datasets (such as Google Books, Gale’s Nineteenth Century Collections Online, the National Library of Australia’s Trove Database, the British Library’s Newspaper Archive) that were not specifically created to enable quantitative analysis. One important—and somewhat limiting—characteristic of our datasets arises from the definitions we employed in our process of extraction. We defined an Australian novelist as any ‘Australian’ author in AustLit who had written at least one novel. As a result, our data is linked to the author, not the novel (and thus cannot indicate titles that received more attention than others). This definition also means that authors are present in the dataset who have written a novel or novels, but who are primarily known—and who received the majority, if not the totality, of critical attention recorded in AustLit—for other literary genres, such as poetry or plays. In other words, these authors we defined as Australian novelists by one set of parameters would not be defined as Australian novelists within literary critical discussion broadly.

To minimise the number of authors in this category, we removed from our datasets authors who had written only one novel. While this step improved the overall focus of the data, it also introduced bias (a dataset absent of first-time authors and excluding one writer in particular, Helen Darville, who received substantial attention—much of it negative—in relation to a single novel). This bias we justified by our interest in tracking the careers of authors; in this respect, first-time novelists cannot usefully be compared to authors with multiple novels. Note that the introduction of bias—while never an aim of quantitative analysis—is only a travesty if data is
perceived as truth or fact. Acknowledging, instead, that all data is value-laden and culturally produced changes the goal from eliminating bias to identifying, and either attempting to accommodate, or make an argument for, the types of bias that inhere in data.

As part of the process of characterising our datasets—to assess the types of analysis and questioning they can support and enable—we created a series of plots for Australian novelists showing the titles they published in relation to the critical commentary they received. Figure 10.2 shows these results for Katharine Susannah Prichard (1883–1969), a politically active Australian journalist and renowned novelist, with the left-hand axis measuring the number of novels published, the right-hand axis the cumulative number of ‘works about’ Prichard, and the x-axis the years of both types of publication.

Figure 10.2. Katharine Susannah Prichard: Number of Novels by and ‘Works About’, 1900 to 2006

According to the results displayed in this graph, Prichard received some critical attention for her first three or four novels, but this attention significantly accelerated in the late 1920s, following the publication of Working Bullocks, Coonardoo, and The Wild Oats of Han. Two further stages of accelerated growth in critical attention follow: first, from the late 1950s to the early 1970s, during which time she published only one novel; and second, from the early 1980s to the mid-2000s, after Prichard’s death. In the 1980s and 1990s, approximately the same number of ‘works about’ Prichard was published as in the previous seven decades. One could easily imagine a whole article exploring these different phases and the factors contributing to the posthumous growth in Prichard’s reputation: Is this an ongoing effect of second-wave feminism, a consequence of the more general vogue for anthologising earlier Australian authors, or something else (for instance, the professionalisation of criticism as ‘research’ in universities in the 1980s, and of research as a key performance indicator in these institutions)?

While the idea of such an article is exciting, comparison of these results for Prichard with the overall number of ‘works about’ extracted, shown in Figure 10.3, raises questions about the
reliability of this type of analysis. Figure 10.3 shows the same stages of growth in critical attention overall—in the 1920s and 1930s, 1960s, and 1980s and 1990s—as Figure 10.2 shows for Prichard. The correspondence between these periods of growth raises the possibility that ‘trends’ in critical attention to Prichard could actually reflect broader shifts in AustLit’s indexing process. Comparison with other authors does not resolve this issue: the same pattern could indicate the relative comprehensiveness of AustLit at particular times, similarities in the critical reputations of those authors, or increased attention to Australian literature broadly; differences could signal dissimilarities in the critical reputations of authors or different degrees of focus, from AustLit, on those individuals. More broadly, the growth in critical attention to Prichard, in these three periods and overall, correspond with what David Carter identifies as three periods of nation-oriented expansion of critical attention to Australian literature: ‘intellectual nationalism’ in the 1920s and 1930s; institutionalisation and professionalisation in the 1950s and 1960s; and ‘neo-nationalism’ in the 1980s and 1990s (2000). There is simply not enough information, from these graphs and AustLit’s account of its scope and content, to determine the significance of these numerical results. Proposing that these trends are reliable because they correlate with the accepted history of Australian literary criticism runs the risk of confirmation bias (of accepting the interpretation that confirms existing, and strongly held, beliefs). These beliefs could equally have shaped the collection of data, with indexing inadvertently focused on those periods when ‘works about’ Australian novelists are perceived to be highest (a process known as selection bias).

Figure 10.3. Total Number of ‘Works About’

Further analysis, however, suggests that indexing has not been overly skewed by such selection bias or by the more comprehensive attention to periodical and newspaper records after 1988 (the only specific information AustLit provides regarding content and scope). Figure 10.4 represents the number of authors per decade who were ranked in the top ten for critical attention each year (a number that must be between ten—if the same ten authors shared the top ten
positions each year—and 100—if the top ten authors were different each year).\textsuperscript{20} If either of the effects described above occurred, we would expect to see a greater range of authors in the periods aligned with the expansion of literary criticism—the 1920s and 1930s, 1960s, and 1980s and 1990s—and a more limited range at other times. As Figure 10.4 shows, this pattern does not occur; instead, there is an overall decline, from the 1920s, in the number of unique authors in the top ten for critical attention.

**Figure 10.4. Number of Unique Authors in the Yearly Top Ten for ‘Works About’, per Decade**

![Bar chart showing number of unique authors in the yearly top ten for 'Works About' per decade](image)

While we believe the data has not been significantly skewed by our extraction process or by AustLit’s indexing practices, and we can thus offer a reliable basis for analysis of trends over time, it should not be understood as either a complete or a random sample of critical attention to Australian novelists. It seems likely the data will have some bias towards well-known authors, so any comparison of those at the top of the critical hierarchy with those lower in the rankings should be conducted with full awareness of this probability. In particular, given the lack of detail from AustLit regarding scope and content, questions relying on numerical accuracy, including those across a range of authors—for instance, the percentage of novelists who were the subject of two ‘works about’—cannot usefully be asked. For those seeking information about individual authors or to compare particular careers, AustLit will be the best resource, supplemented by intensive checking of records to ensure that all ‘works about’ are discovered.

The datasets we have produced—and which provide the basis for the second part of this paper—include all the data extracted, as well as separate files for the number of ‘works about’ received by the top 50 authors overall from 1900 to 2006, and in three categories of publication from 1950 to 2006 (academic journals, literary magazines, and newspapers).\textsuperscript{21} These top 50 datasets are created from the total extracted data. However, we have published them separately, anticipating that most literary scholars will be particularly interested in the authors who receive the most ‘works about’, but may be unable to parse the data to produce these results themselves.
Canons
The process of literary reputation making has been explored by Pierre Bourdieu, in his groundbreaking sociological work on the literary field (1983; 1996), and by scholars in both literary studies and empirical literary studies. Although Bourdieu is frequently cited as a reference point for this latter group, unlike Bourdieu—who focuses on the ‘struggle’ among multiple groups, institutions, and individuals in the literary field—researchers in empirical literary studies are generally united in their censure of literary criticism as an activity that does not and cannot do what it supposedly claims to do: gauge literary quality and assign authors positions in a hierarchy on that basis. According to Marc Verboord, lacking a ‘clear theoretical and empirical grounding’—and especially without ‘unequivocal standard[s] to measure literary quality’—literary criticism is ‘methodologically unsustainable’ (2003, 259). Similarly, C. J. Van Rees contends that, ‘most of the critical tenets on a writer’s work are defective insofar as they fail to be empirically testable’, and as a result, literary critics ‘lack the capacity of ascertaining the intrinsic quality of a work of art’ (1987, 275, 276). Such statements clearly demonstrate the same uncritical positivism, theoretical naivety, and misunderstanding of the hermeneutics of literary criticism that Ramsay identifies in computational stylistics, and show why these studies have had no effect on mainstream literary criticism discussion of reputation.

These features of their approach also underpin the erroneous view of empirical literary scholars that the findings of their studies—that factors external to the text affect the likelihood of an author receiving critical attention—challenge the very grounds of literary criticism. In fact, while the Romantic ideal of the author as solitary genius remains an influence in literary criticism, it would be difficult to find a literary scholar who did not recognise that value judgements are at least partly culturally produced—that is, the product of social, economic, national, institutional, and professional interests. Indeed, although empirical literary scholars are apparently unaware of them, one of the things that literary critical analyses of reputations—which generally take the form of critical studies of particular authors, including Nathaniel Hawthorne (Thompkins 1983) and, in the Australian context, Peter Carey (Turner 1993), Helen Garner (Darcy 1999), Patrick White (During 1996), and Tim Winton (Dixon 2005)—demonstrate so well is how various institutional, political, economic, and social factors contribute, at different times and in different ways, to the construction and discussion of these authors.

Our analysis of critical attention to Australian novelists aims to do something different from either group. Like the literary critics, we accept that literary reputations are the outcome of complex, ongoing processes involving multiple factors. However, while the potential certainly is present in the data to identify and explore the hierarchical valuation of particular authors, we are concerned instead with literary reputation in the aggregate. Like empirical literary scholars we take a data-rich approach, but we are not interested in measuring the extent to which different factors—of ability, personal and institutional affiliation, and history and culture—contribute to the reputations of novelists; nor do we intend this study to disprove literary critical assessments. Instead, we explore broad trends in the critical field, including changes in the distribution of critical attention over time and in the relationship between different forms of literary criticism, as well as the role of gender in these shifts. The results we discuss in what follows, in other words, are as much concerned with the historiography of Australian literary criticism as with the reputations of authors in that field.

One widely accepted trend in literary criticism since the late 1960s and 1970s—motivated, in particular, by identity-based political movements such as Marxism, feminism, and postcolonialism—is the widening of the literary canon, with the traditional great author revealed as a white man and forced to cede ground to a range of other groups and perspectives. Given this
well-established trajectory, the results in Figure 10.4 are surprising, as they show an overall decline, in each decade since the mid-20th century, in the number of unique authors who are ranked in the yearly top ten for critical attention. Quite in contrast to the view that the canon is expanding, this result suggests—in relation to Australian novelists at least—a narrowing of critical attention. For example, in the 1990s, only 36 authors appeared in the yearly top ten across the decade, compared with 84 in the 1920s.

Figure 10.5 provides a useful context for, and explanation of, this seemingly contradictory trend. This graph shows the proportion of critical attention accorded to the top ten authors per decade, revealing an overall decline from the mid-20th century to the 2000s (notwithstanding a slight increase in the 1960s and 1970s). In other words, despite greater agreement on the most important Australian novelists as the century progressed (see Figure 10.4), critical attention spread to an increasingly large group of authors. Taken together, the results in these two graphs suggest growing critical confidence, with increasing agreement on the best authors signifying consensus on what constitutes a ‘good’ or ‘great’ Australian novel, at the same time as the spread of critical attention signals confidence in the overall quality—and worthiness for critical attention—of the range of novelists.

**Figure 10.5. Proportion of Critical Attention to Top Ten Authors, per Decade**

A different picture of this confidence emerges if we consider particular categories of criticism: specifically, academic journals and newspapers. In both cases, the concentration of critical attention to the top ten authors is greater than for overall results, especially from the 1970s, as Table 10.2 shows. While we might expect academic criticism to be more focused on a select group of novelists than overall criticism, it is interesting that this relatively concentrated focus also occurs in newspapers, where publishers compete—commercially, not simply at the level of cultural value—to have their authors featured. Bourdieu’s description of the function of critical pronouncements—‘Every critic declares not only his judgment of the work but also his claim to the right to talk about it and judge it’ (1983, 317)—suggests a reason for the relative
concentration of critical attention in both areas. To maintain a reputation and capacity to offer opinions, a critic cannot afford for his or her judgement to differ, repeatedly, from the group: that is, from the consensus about what is good literature. In the worlds of newspaper and academic criticism, where having the right to judge brings a range of professional and social rewards, it makes sense that critical consensus would be relatively strong.

Table 2: Proportion of Critical Attention to Top Ten Authors Per Decade

<table>
<thead>
<tr>
<th>Decade</th>
<th>Overall</th>
<th>Newspapers</th>
<th>Academic Journals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1950s</td>
<td>29%</td>
<td>45%</td>
<td>33%</td>
</tr>
<tr>
<td>1960s</td>
<td>36%</td>
<td>36%</td>
<td>36%</td>
</tr>
<tr>
<td>1970s</td>
<td>33%</td>
<td>40%</td>
<td>39%</td>
</tr>
<tr>
<td>1980s</td>
<td>20%</td>
<td>28%</td>
<td>34%</td>
</tr>
<tr>
<td>1990s</td>
<td>15%</td>
<td>30%</td>
<td>32%</td>
</tr>
<tr>
<td>2000s</td>
<td>12%</td>
<td>25%</td>
<td>32%</td>
</tr>
</tbody>
</table>

Figure 10.6. Percentage of Authors Who Feature in the Top 50 for Both Academic and Newspaper Criticism, 1950–2006

While the consensus within academic and newspaper criticism is relatively high, there is declining agreement, between the two areas, on the authors most worthy of discussion. Figure 10.6 depicts the proportion of Australian novelists who appear in the top 50 for both academic and newspaper discussion from the 1950s, and shows—with the exception of the 1980s—a marked decline in this crossover to the 2000s. Whereas 63 per cent of Australian novelists in the top 50 for academic and newspaper criticism appear in both categories in the 1950s, by the 2000s this figure falls to 21 per cent. The exception of the 1980s provides a framework for explaining the broader trend. Following a decade of intensive cultural nationalist funding, and centring on Australia’s 1988 bicentenary, the decade of the 1980s is widely recognised as an exception period
in Australian cultural history, when a broad set of concerns—regarding feminism, Aboriginal land rights, multiculturalism, and national identity—animated both public discussion and academic discourse, and were expressed with particular force and impact in Australian fiction, especially novels (Davis 2009; McPhee 2001).

The fact that the 1980s are recognised as exceptional—specifically in the connection that occurred between academic and public discussion of literature—tallies with Graeme Turner’s description of a broader international shift in the public sphere since the 1960s, marked by the growing alienation of ‘the academic study of literature . . . from the language of public debate’ (1999, 9). In this light, the results displayed in Figure 10.6 provide an interesting perspective on what we have called a trend of growing confidence in pronouncements about Australian literature. Although academic critics may increasingly agree on the Australian novelists who are worthy of discussion, the distance of these pronouncements from public discussion—brought into focus by the very small readership of academic journals compared to newspapers—suggests the cultural authority exercised by this academic agreement, and its weight in directing broader public discussion, has significantly declined.

We conclude by touching briefly upon a trend—in gender—that highlights one important consequence of this separation of academic and newspaper discussion of Australian literature. Figures 10.7 and 10.8 show the proportion of women among the top 10, 20, and 50 ranked authors per decade in academic journals and newspapers, respectively. The fact that approximately half of the top 10 and 20 most-discussed Australian novelists in academic journals in the 1990s and 2000s are women suggests that the feminist deconstruction of the canon continues to influence conceptions of literature in this area.

**Figure 10.7. Proportion of Women Authors in the Top 10, 20, and 50 for Academic Attention, 1950s–2000s**
In newspaper criticism, by contrast, only 10 to 20 per cent of authors in the top 10 and 20 rankings in these two decades are women. The significant difference in gender trends in the canon (or canons) of Australian literature suggests that the equation of great author and man has recently reasserted itself—indeed, has become more trenchant than in the 1950s and 1960s—in the public sphere. To avoid losing the progress that feminist critics made in challenging male-oriented definitions and valuations of literature requires, as a necessary first step, a reengagement of academic critics in public discussion of Australian literature.

Figure 10.8. Proportion of Women Authors in the Top 10, 20, and 50 for Newspaper Attention, 1950s–2000s

All of these interpretations of the data are contestable; indeed, they can be contested through further exploration of the datasets we have published. A major aim of this paper has been to encourage and enable such contestation and future analysis by providing characterised and reliable data for researchers to use. The work of gathering, checking, and analysing this data to understand the types of questioning it can enable and support is a significant task, requiring expertise that is increasingly located in digital humanities. Drawing on astronomy, we have proposed a mode of publication that enables this technical and methodological work to be used and recognised by the wider scholarly community, while contributing to the broader humanities endeavour of raising new questions, offering new knowledge, and contributing to public debate. Ultimately, the technical and methodological cannot be separated from the critical and theoretical. Characterising—that is, considering the epistemological implications—of data is an explicitly theoretical and critical process. Equally, humanities research that draws upon digital resources and data requires deep understanding of the methodological processes and decisions underpinning those constructs. Of necessity, digital humanities must be an arena where these two fundamentally intertwined aspects of digital and data-rich research are combined and balanced.
Works Cited


Notes


2 The fact that only one of the five existing Stanford Literary Lab pamphlets includes access to data, despite all being published online, shows that this lack of openness about data is more than an effect of publishing strictures.

3 For example, the data release papers for the massive Sloan Digital Sky Survey underpinned a decade of scientific work. They have provided the basis for hundreds of scientific papers and been cited many thousands of times (see Stoughton et al. 2002).

4 This topic was discussed in respect to tenure in a special issue of the *Journal of Digital Humanities* (see, for example, Cavanagh 2012; Nowviskie 2012).


8 The number of ‘works about’ an author does not indicate different levels of critical attention: one instance is recorded whether the author is the subject of an entire monograph or critical essay, or is just mentioned in passing in a newspaper article or academic review. While
collapsing one type of variation in the critical field, the data captures the essence of literary reputation better than, say, a word count, because it brings into sharp relief—indeed, amplifies—the presence of those authors who emerge as reference points for critical discussion.

9 ‘Non-AustLit’ are usually well-known writers (such as George Eliot, Harriet Beecher Stowe, and Charles Dickens) included in AustLit because they are deemed to have profoundly influenced particular Australian authors. The list of authors used in this automatic extraction process also excluded non-Australian authors in the ‘banned books’ AustLit subset, which lists works censored in Australia.

10 ‘Other’ incorporates 23 categories of publication, each with less than 1 per cent of total ‘works about’, including prose, obituaries, poetry, essays, theses, and bibliographies.

11 For example, a pseudonym is indicated by ‘Writing name for’ as well as ‘Used as a writing name by’.

12 Each point on the plot is a single author, and only authors with four or more novels are shown. The majority of authors are clustered in the bottom left, with a moderate number of novels and critical works. The extreme points in the graph are pulp fiction author Grover Marshall on the far right (with 683 published novels and 3 ‘works about’), and literary fiction author David Malouf on the upper left (with 6 novels and 1,103 ‘works about’).


15 AustLit, ‘About Scope’.


17 For discussion of the difficulties involved in using the Google Books archive and its Ngram viewer, see Cohen 2010; Jockers 2010.

18 As AustLit explains, ‘The Hand that Signed the Paper’ purported to be a family history based on oral testimony, relating a family’s role in Jewish concentration camps’ (AustLit, ‘Darville, Helen’, http://www.austlit.edu.au). The book won a number of prestigious awards—including the Vogel Prize, the Australian Literature Society Gold Medal, and the Miles Franklin Award—and provoked enormous controversy when Darville’s actual background was revealed.

19 Figure 10.3 is also one of the diagnostic plots we used to explore and test the process of extraction and the resulting dataset. The obvious peak in ‘works about’ in 1924—which might indicate an error or a focus by a specialist dataset—actually relates to the publication of The Boy in the Bush by D. H. Lawrence and M. L. Skinner. Due to Lawrence’s international fame, this work attracted considerable critical attention in Australia and elsewhere.

20 This number could actually be greater than 100 if more than ten authors were in the top ten in any year (for instance, if tenth place was shared by two or more authors receiving the same amount of critical attention). The nature of the data produces some bias in the results: because there was much less critical attention to Australian novelists early in the 20th century, multiple authors were more likely to tie on the number of critical works. By decade, the average number of authors in the yearly top ten is: 1900s, 12.9; 1910s, 15.4; 1920s, 16.9; 1930s, 12.4; 1940s, 11.9; 1950s, 12.3; 1960s, 10.7; 1970s, 10.8; 1980s, 10.4; 1990s, 10.1; 2000s, 10.6.

21 These periodicals were manually categorised. Academic journals were identified retrospectively, based on those that employed peer review in 2007.

22 This field comprises predominantly sociologists and psychologists who use empirical methods to investigate aspects of literary culture, including reception and audience studies (see, for example, Janssen and van Dijk 1998; Zyngier et al. 2008).
For instance, research in empirical literary studies shows that critically acclaimed authors are more likely to have pursued academic studies in literature and be members of writers’ organizations (Gerhards and Anheier 1989); that critics are more likely to pay attention to authors who have already received critical recognition (Verdaasdonk 1987, 238–9) or are involved in public roles besides writing (Janssen 1998); and that the reputation of an author’s publishing house is often decisive in determining which authors will gain critical attention (Nooy 1991, 509–10; Van Rees and Vermunt 1996; Janssen 1997; Nooy 2002).

An exception to these studies in literary criticism of individual authors is David Damrosch’s analysis of what he terms the ‘hypercanon’, ‘countercanon’, and ‘shadow canon’ of Romantic British and world literature (2006).

In three decades, there were more than ten authors in the top ten: in the 1900s and 1910s, 11 authors were in the top ten; in the 1960s, there were 12.

For a detailed discussion of the relationship between gender trends in newspaper and academic criticism and the feminist movement, see Bode 2012a, 131–67.