Just as digital technologies have transformed publishing, so too are they transforming publishing studies. As in other disciplines, such as archaeology, by changing the methods and scale of research, digital resources and approaches enable new perspectives on and insights into the history of publishing. With this digital transformation in mind, the chapter’s first part identifies three principles that I think are essential to bear in mind when conducting publishing research with large digital collections. In the third and final part, I explore another issue relevant to digital research—data publishing—that has the potential to redirect and invigorate the future of publishing studies, including by connecting it to new collaborators and publics.

The chapter’s second part demonstrates these principles and possibilities with reference to previously unknown features of early Australian publishing: of fiction in nineteenth-century newspapers. These periodicals are the “large, vigorous and thriving” enterprises of my title (Morrison 1998, 308), and publishing historians have long known that some of them published fiction. More particularly, they have argued that most of this fiction appeared in major metropolitan newspapers, especially the weekly companions to the major metropolitan dailies—including Melbourne’s Leader, Sydney’s Australian Town and Country Journal, and Brisbane’s Queenslander—and that most of these stories (around 80 percent by one estimate) were by British authors (Morrison 1998, 315). This understanding has contributed to a (perhaps the) key legacy of Australian publishing studies: the notion that Australian publishers—and hence readers—have for most of their history been dominated by British interests, publishers and authors. Summarising this widely held view in the Cambridge Companion to Australian Literature, Elizabeth Webby wrote that, “for much of the nineteenth century and indeed afterwards, Australian readers were mainly interested in books by English authors” (2000, 50).

Applying new digital methods to new digital resources—namely, the National Library of Australia’s (NLA) extensive mass-digitised collection of historical Australian newspapers—enables a very different version of Australian publishing, literary and reading history. Based on data drawn from Trove Newspapers and published in To be continued: The Australian Newspaper Fiction Database (National Library of Australia nd.; Bode and Hetherington 2017), I show that nineteenth-century Australian newspapers were undoubtedly the most successful and prolific of Australia’s early independent publishers. Not only was fiction publishing much more widespread and actively pursued in Australian newspapers in the nineteenth century than has been recognised, it was cosmopolitan from its origins, encompassing fiction from around the world. At the same time, local writing had a much greater presence, more organised systems of distribution, and a more receptive and substantial market, than literary and publishing historians have understood. From the time Australian newspapers began routinely to publish fiction (the mid-1860s), early Australian fiction publishing was both enmeshed in globalised cultural and economic networks and distinctively local, adaptive and independent.
Methods and principles for digital publishing studies

From 2013 to 2016 Carol Hetherington and I worked to identify and explore fiction in *Trove’s* digitised newspapers. While most Australian researchers are familiar with *Trove* and its newspaper collection, some—perhaps many—are unaware that this is the largest open-access collection of digitised historical newspapers in the world, with significantly more searchable pages than related collections including *Chronicling America, Europeana or Papers Past*. Of course, just because a collection is big does not make it complete; nor does it mean the documents it contains are representative of those that once existed.

By comparing the newspapers digitised by *Trove* with historical records, such as indexes of newspapers created for advertisers, I estimate that, at the time we stopped harvesting from *Trove* (July 2015), 28 percent of nineteenth-century Australian newspapers had been digitised. This percentage is perhaps lower than a lot of people assume when conducting historical research with *Trove*, and certain parts of the collection are also over- or under-represented. For instance, newspapers from Queensland and Western Australia have been digitised at a higher rate than for other colonies, while provincial newspapers are generally underrepresented. This recognition that archives and other collections are partial and reflect a particular version of history—created by collection practices, contingent notions of value and meaning, and economic and political considerations—is one that publishing studies has always understood and emphasised, and it should not be minimised or obscured when the research shifts from analogue to digital collections, regardless of how extensive the latter may be.

My first principle for conducting publishing studies with digital technologies is, thus, an old one: namely, that the collections we work in have histories. We need to know those histories—including the effects of our own engagements with those records—to understand the relationship between the collections we investigate and the publishing context we seek to understand.

After establishing what proportion of historical newspapers have been digitised, and areas of under- and over-representation, the task was to find the fiction these periodicals contained. Rather than applying complex text mining and machine learning algorithm, as the ‘Viral Texts’ project does to analyse reprinting in historical American newspapers (Cordell and Smith nd; Smith et al. 2015), our analysis of *Trove* began with arguably the most familiar of all components of digital research infrastructure: the search box. However, we used it in a different way to how keyword searching is typically approached in the humanities.

What we did not do was to enter the names or titles of known nineteenth-century authors. The limitations of that approach are (at least) four-fold. First, it would produce lots of irrelevant results, including literary gossip, obituaries and book reviews. Second, it would be very time consuming, as it would require the researcher to work through bibliographies of fiction and enter hundreds or even thousands of authors or titles into *Trove*. Third, this approach would be self-confirming, in that it would yield only the authors and titles the researcher expected to be present. Even if this expectation comprised all known authors and titles, as would be the case working with bibliographies, such an approach would inevitably reinforce existing knowledge of fiction. Finally, this approach is ahistorical in that it fails to accommodate the common changes in title for, and especially the widespread anonymous and pseudonymous publication of, fiction in early Australian newspapers.

The solution devised in response to these practical and conceptual problems is one that I call the “paratextual method.” Paratext, of course, refers to elements—here, textual—that surround, signal and introduce publication events. In nineteenth-century newspapers, as today, paratextual features are very consistent, enabling readers to open the pages and differentiate, at a glance, the advertising from the feature articles, the letters to the editor from the crime reporting, and the sports reporting from the fiction. The paratextual method adapts this mechanism—or technology—for supporting human searching and reading to digital discovery. Thus, into the search box I entered not authors or titles, but words that consistently appeared in the paratexts of fiction in early Australian newspapers, such as “serial,” “story,” “our author,” “tales and sketches,” “our novelist,” and (most successfully) “chapter.”
Because of the way Trove’s relevance-ranking algorithm works to privilege articles where words appear in the title, and especially when they recur throughout (as with “chapter”), the vast majority of the initial tens of thousands of results for searches with these terms were fiction. We then used Trove’s Application Programming Interface (or API—a set of protocols for interacting with a computer system, in this case a database) to export the results in large batches. For each search term, this approach was employed until a substantial number of irrelevant records appeared in results: a sign that a particular paratextual term, in its interaction with Trove at a specific point in time, had exhausted its usefulness. (For each term harvesting was performed multiple times over a two-year period, yielding new results as additional newspapers were digitized.)

The “paratextual method” embodies a second principle for conducting publishing studies with large digital collections: the need for methodologies that articulate between disciplinary knowledge and the affordances and systems of the digital infrastructure available. In this case, a focus on paratextual search terms reflects an understanding of the newspaper genre, and of fiction publishing in nineteenth-century periodicals. But it can only be systematized—that is, used to investigate a large digital collection in a reliable and consistent way—by virtue of Trove’s specific digital affordances: its segmentation of digitized pages into articles; its manual correction of the Optical Character Recognition [OCR] text in article titles; the nature of its relevance ranking algorithm; and the provision of an API to enable mass searching and harvesting of results.

This analysis of digitised newspapers yielded over 200,000 article records. But it was not simply a matter of extracting the data and beginning analysis, because the format in which the API provides information is not particularly useful for publishing research. Figure 1 is an example of the html records that were extracted, along with associated text files. This example contains a wealth of information in a single field “<heading>,” including the story’s title (“WINNUM WHITE’S WICKEDNESS”) and information about copyright (“The right of publishing this Novel in the Northern District has been purchased by the Proprietor of the ARMIDALE CHRONICLE”) as well as the chapter number and chapter title published in that instalment. Other “<heading>” fields
arising from the same “chapter” search contained much more information than this one, including other periodicals in which the work had been published, descriptions of the story (for instance, as “sensational,” “original” or “tragic”) or signature details (that this story is “by the author of …” other titles). In other records, the “<heading>” field contained only the chapter number.

These differences in the information presented in particular fields, as well as the inclusion of irrelevant results, meant that significant manual data processing was required to transform what was extracted from *Trove* into data useful for research. Different instalments of stories also had to be collected under a single title, and Carol also conducted extensive bibliographical research on these records, for instance, to identify changes in the titles of stories or the origins of anonymously and pseudonymously published works. The third and final principle for conducting publishing studies with large digital collections, evidenced by the above example, is that digital research is not the quick or lazy option. Although we hear a lot about how digital resources radically increase access to information, the form in which that information is available is often not one that suits academic research. Extensive work, and once again, significant disciplinary expertise, goes into constructing useful data. As well as taking an enormous amount of time, this process adds considerable value to the resulting data—an issue I will return to with respect to data publishing, in the final part of the chapter.

Based on this analysis of *Trove* digitized newspapers, we discovered over 21,000 novels, novellas and short stories, including unique publications and republications, in nineteenth- and early-twentieth-century Australian newspapers. Other details of that dataset, available through the *To be continued* database, include:

- 16,648 fictional works published in the nineteenth-century, made up of 9,249 works of extended fiction (stories serialised over two or more newspaper issues or amounting to 10,000 words or more in a single issue)\(^2\) as well as 7,399 short stories;
- 4,515 fictional works in early twentieth-century newspapers (published from the start of the twentieth century until 1914), made up of 3,507 works of extended fiction and 1,008 short stories;
- 4,146 author names;
- 5,671 titles where an author’s name is not given and, even after extensive bibliographical research, we have not been able to confirm an identity; and
- 194,108 individual text files (some of these are stories in their own right; a great many comprise instalments that needed to be grouped with other text files).\(^3\)

When one considers that, prior to this analysis, the bibliographical record for fiction in nineteenth-century Australian newspapers numbered in the hundreds, this discovery of almost 17,000 titles represents a significant increase on existing knowledge and confirms the importance of newspapers as sources of fiction for colonial readers.

**New discoveries about early Australian publishing**

In what follows I focus on fiction published in the nineteenth century, especially between 1865 and 1899, when 98 percent of the titles discovered in this period appeared. More specifically still, I concentrated on extended fiction, as defined above, because these titles imply a particular mode of publishing and reading. Where stories completed in a single newspaper issue suggest incidental publishing and reading—with such content often selected simply to fill column inches, and likely read in a casual manner—extended fiction required deliberate sourcing and publishing by editors, and implies more intensive or committed engagement from readers. This subset is in itself substantial—over 9,200 titles—and exploring it makes possible four findings that substantially refigure understandings of publishing in Australia in the nineteenth century, and of the transnational circulation of fiction in this period.

As already noted, publishing historians have recognized nineteenth-century Australian newspapers as fiction publishers for many decades, with the understanding that most of this fiction was British. Among the fiction harvested from *Trove*, British works were prevalent, comprising about
50 percent of titles where authors' identities were discovered. But the idea that British fiction was essentially all that was published is far from true.

The first key finding from this project is that fiction in nineteenth-century Australian newspapers originated from a much greater and more diverse range of national and proto-national contexts than has been understood. American fiction was widely published, representing around 20 percent of the titles by known authors discovered, and there was also a significant number of French and German titles in translation as well as fiction from other British dominions (Canada, New Zealand, South Africa) and even further afield, including from Austria, Holland, Hungary, Italy, Japan, Russia and Sweden. Much of this fiction was popular, including a substantial amount of British sensation and adventure fiction and American dime stories, including crimes, romances and westerns. But a surprising number of titles were by literary, even canonical writers: Honoré de Balzac, Charles Dickens, Benjamin Disraeli, Alexandre Dumas, George Eliot, Thomas Hardy, Victor Hugo, Henry James, Harriet Beecher Stowe, Eugene Sue, William Thackery, Anthony Trollope, Ivan Turgenev, Mark Twain, Oscar Wilde and Émile Zola are all represented in the To be continued database. As with the sheer scale of the fiction published in newspapers, this list of literary writers indicates how different this nineteenth-century reading context was from the one we associate with newspapers today.

Alongside this wide range of international writing was a significant amount of Australian fiction: both known and new. Some notable findings include previously unlisted fiction by Catherine Martin and Jessie Mabel Waterhouse; a new Australian author, John Silvester Nottage, responsible for multiple full-length novels; and new titles by “Captain Lacie” and “Ivan Dexter,” in addition to the discovery that both were well-developed pseudonyms for James Joseph Wright. Indeed, although British author of sensation fiction, Mary Elizabeth Braddon, has been identified as the most widely-published author in the Australian colonies, based on the fiction harvested from Trove, Wright—a local author—holds this position. Certainly, he emerges from this analysis as one of Australia’s most prolific writers. More significant to understanding the publishing context than these individual literary discoveries is the scale of Australian fiction present. Such stories make up around 25 percent of titles by known authors discovered in this project, with many hundreds of other works by unknown authors indicating—indeed highlighting—an Australian origin.

This brings me to a second key finding arising from analysis of this extensive new collection of fiction: not only was there more local fiction in nineteenth-century Australian newspapers than has been recognized, it was also accorded significantly more value. This statement might give readers pause: how can I know that readers valued Australian fiction? Local stories could have been published simply because they were cheaper or more readily available than the imported product? However, the paratextual information surrounding these stories supports this statement in providing insights into the meanings and values informing both publication and reception. Editors aim to sell newspapers—and fiction was one of the main ways they achieved this aim. So it makes sense that they would foreground features of the fiction that they thought readers would look favourably upon. And in nineteenth-century Australian newspapers, it was routinely the case that local origins of fiction were highlighted.

Some individual examples will serve to give a sense of this local emphasis. Thus, in the 1890s, the Jewish Herald assured readers that “A Jewish Wife” was “An Australian Story by an Australian Author” (#13201); and the Traralgon Record described “The Dis-Honourable” as “A newly published and fascinating story of Australian life” (#6074). Such emphasis was not limited to the 1890s. In 1887, the Horsham Times was one of many newspapers to present “ALMA” as “An Original Australian Story” by a “New South Wales” member of parliament (#4945); and in 1875 in the Sydney Mail, the first chapter title for “Investing Uncle Ben’s Legacy; or, Mr. McTailing’s Mining and Matrimonial Speculations. A Tale of the Mining Era” set the scene firmly in the colonies with: “A failure in Melbourne and a fresh start in Sydney” (#14877). The local origins of this latter story are further emphasized by its attribution to the very Australian-sounded “Old Boomerang,” a pseudonym for journalist J. R. Houlding, who emerges from this analysis as the most widely published colonial author of the 1860s and 1870s.
Hundreds of stories simply featured Australia—or places therein—in their titles. Some examples by well-known authors include J. D. Hennessey’s “The Bells of Sydney” in 1896 (#6064; #6065), Lilian Turner’s “By the Blue Australian Mountains” in 1894 (#7820), and Rolf Boldrewood’s “The Wild Australian” in 1877 (#7664). Lesser known or unknown authors did the same, including V. L. Thomas’s “An Australian Anarchist” in 1896 (#15438), Ralph De Peveril’s “An Australian Settler’s Tale” in 1874 (#149; #150), and “Bush Life in Australia: An Original Tale,” published anonymously in 1859 (#423). Indeed, so consistently were local origins emphasized that, although Australian titles make up only 25 percent of those by known authors, “Australia/n” was the third most common word in titles and subtitles across all extended nineteenth-century fiction, after “story” and “tale.” Some overseas fiction—particularly from America—was even adapted to make it seem as if it came from Australia. Thus, “Florabel’s Lover, or, Rival Belles” by American writer Laura Jean Libbey became “The Rival Belles of Parramatta” in multiple provincial newspapers in 1888 (#3042; #3043; #3044; #3045; #3561; #14589; #14590; #14591; #14592; #14593); and Old Sleuth’s “The American Detective in Russia” was serialized in various sites as “Barnes, the Australian Detective” (#6415; #6416; #6417; #6418; #6419; #6420; #6421).

The data represented in Figure 2 provides a broader sense of this foregrounding of local origins. It shows the place with which stories were identified (by titles, by-lines, copyright information and other prominent paratextual and textual features) where the actual nationality of authors is known. For all national or proto-national categories (American, Australian, British and other) the presentation or inscription of fiction was more likely to align with the actual origins of the author than otherwise. But there are significant differences between these categories. American and other fiction was much more likely to be depicted as coming from elsewhere than either British or Australian writing. By contrast, editors were equally likely to represent British and Australian fiction as coming from Britain or Australia, respectively. This trend suggests that British fiction was seen as desirable and of interest to readers; but not necessarily more so than Australian writing.

Figure 2: Proportion of titles by authors of known nationalities as nationality was inscribed
The cultural value accorded to Australian writing is reinforced by the attribution of fiction. If we allow that an editor was more likely to attribute a title to a particular author if that author was seen as prestigious or noteworthy, then trends in attribution can also be used to explore editorial assumptions about colonial readers’ interests. In this light, it is significant that British fiction was more likely than most other fiction to be attributed to a named author in nineteenth-century Australian newspapers; but local fiction, especially by men, was the most likely to be attributed. Taken together such foregrounding of local origins and authors suggests that Australian fiction not only appeared much more frequently in nineteenth-century Australian newspapers than has been recognized; its publication was also foregrounded as interesting and valuable to readers.

The third major finding from this project is, simply put, a new structure and organization for nineteenth-century Australian publishing and literary culture. Scholars have focused on metropolitan newspapers as fiction publishers, assuming that fiction was rare—if present at all—in provincial periodicals. I have found, to the contrary, that provincial newspapers published more fiction than their metropolitan counterparts: 55 percent of the extended nineteenth-century fiction identified. This proportion is made more remarkable by the facts that provincial newspapers are underrepresented in Trove, and only started publishing fiction in earnest in the 1880s, whereas metropolitan newspapers began in earnest in the 1860s. Thus, the metropolitan newspapers that have received almost all the critical attention emerge as less prolific in their publication of fiction than their neglected provincial counterparts.

More significantly for Australian publishing history, metropolitan newspapers also emerge as less interconnected than provincial ones in how they sourced this fiction. Among the titles identified in Trove’s digitized newspapers, it is not uncommon to find two or three, sometimes even four, metropolitan newspapers publishing the same overseas story around the same time. Not infrequently, tracing the publishing histories of these works confirms that they were syndicated by British or American companies and suggests that metropolitan newspapers strategically banded together to purchase particular international stories, or purchased stories individually and sold them on to other colonial newspapers. In contrast to these occasional metropolitan groupings, large clusters of provincial newspapers published many of the same titles, around the same time, and often in an identical sequence.

![Network graphs of shared publications](image)

1882-1885

<table>
<thead>
<tr>
<th>Eastern Ensign</th>
<th>Nepean Times</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>border Watch</em></td>
<td><em>Singleton Argus</em></td>
</tr>
<tr>
<td>Kapunda Herald</td>
<td><em>Adelaide News</em></td>
</tr>
<tr>
<td>Gympie Times</td>
<td><em>Riverina Recorder</em></td>
</tr>
<tr>
<td>Port Adelaide News</td>
<td><em>Warwick</em></td>
</tr>
<tr>
<td>Horsham Times</td>
<td><em>Shoalhaven Telegraph</em></td>
</tr>
</tbody>
</table>

1885-1888

<table>
<thead>
<tr>
<th>Mackay Mercury</th>
<th>Border Watch</th>
</tr>
</thead>
<tbody>
<tr>
<td>Queensland Times</td>
<td><em>Singleton Argus</em></td>
</tr>
<tr>
<td>Ipswich Herald</td>
<td><em>Gympie Times</em></td>
</tr>
<tr>
<td>Shoalhaven Telegraph</td>
<td><em>Riverina Grazer</em></td>
</tr>
<tr>
<td>Northern Ensign</td>
<td><em>Warwick</em></td>
</tr>
<tr>
<td>Warwick Argus</td>
<td><em>Gympie Times</em></td>
</tr>
<tr>
<td>Port Adelaide News</td>
<td><em>Shoalhaven Telegraph</em></td>
</tr>
</tbody>
</table>

Figure 3. Excerpts from network graphs of shared publications, 1882 to 1885 and 1885 to 1888
Figure 3 shows two examples of these groupings, in the form of excerpts from network graphs for different periods. In these graphs, a connection (or to use network terminology, an edge) between two newspapers indicates that they published a title or titles in common, with the thickness of the line growing as the number of common titles increases. In this force-directed graph, the distance between different newspapers also relates to publishing patterns: specifically, the algorithm forces individual publications (or nodes) closer together the more they share links in common with other newspapers.

Based on these publishing patterns, and after detailed examination of publishing histories and newspaper pages, we have identified eleven separate newspaper fiction syndicates operating in the Australian provincial press in the latter part of the nineteenth century. Only one of these had been known to publishing historians previously (Morrison 2005, 210–12, 253–56). These syndicates operated in many cases across multiple colonies and could include seventy or more shared titles, and forty or more individual newspapers; and that is from a collection where under 30 percent of newspapers are digitized, and where provincial newspapers are underrepresented. In all probability, therefore, these syndicates were much more extensive than we have been able to discern.

Significantly, when the circulations of these provincial newspapers are combined, the readership of many of these syndicated stories would have been as large, if not considerably larger, than for fiction published in metropolitan venues.

Some of these syndicates published mostly international fiction, and may have been international enterprises with distribution networks in Australia or colonial enterprises that sourced their fiction primarily from overseas. But others published mostly Australian fiction, and at a scale that makes them—even individually—the most significant publishers of Australian writing in the nineteenth century, and probably up until the latter part of the twentieth century. For instance, in just four years, between 1880 and 1884, one of these syndicates serialised twenty-six Australian novels, plus an additional two that are almost certainly Australian. By comparison, AustLit records indicate that the most prolific local book publisher of this period—George Robertson—in thirty years (from 1860 to 1899) published only nine Australian novels (Bode 2012, 44). Put simply, the prevalence of fiction in provincial newspapers and the operations of these syndicates indicate that publishing historians have been looking in the wrong place—to book publishers and metropolitan newspapers—when seeking to understand fiction publishing in the Australian colonies.

The fourth and final finding relates to the contents of these fictions, particularly the Australian titles. When considering thousands of novels and novellas, one cannot just read them: it would take decades. But digital methods offer new ways to explore and understand large collections of text. The approach I used combines a common method in digital literary studies, topic modelling (see for example Goldstone and Underwood 2014), with a method fairly widespread in other disciplines (such as ecology and economics) but new to the analysis of literary works: decision trees. Both use machine learning algorithms to sort data. The first sorts words from a large corpus into groups based on the likelihood that these words will appear in the same documents. The second identifies which of these word clusters or topics are most likely to appear in or to characterise fiction of a specific category (in this case, author nationality).

I like this method because it aligns corpus-wide topics—lexically contextualised but separate from the literary works these patterns arise from—with historical categories of documents. Although the calculations are complex, this integrated approach resonates with a widespread understanding in literary studies of the way in which fiction and social forms relate: namely, that authors in particular social groups are likely to demonstrate certain literary tendencies. Showing that certain groups of words or topics tend to occur in women’s fiction, for instance, does not mean that all women writers use these words, or that in all the women’s writing that displays this characteristic, the words appear with the same frequency. Rather, in indicating tendencies or inclinations in the type of literary language used by groups of authors, this integrated method employs quantitative methods without reducing literary meaning to them. Figure 4 shows the topic—or the most prominent 200 words in the topic—identified by this method as the most likely to characterise Australian fiction in nineteenth-century Australian newspapers.
The prominent words in this topic suggest what is commonly known as the bush tradition in Australian fiction: cattle, horses, creek, sheep, kangaroo, men. But other prominent words suggest a different bush tradition to the one we know: I am referring, here, to words such as township, veranda, buggy, homestead, station, on the one hand, and black, blacks, mob, tracks, on the other. Reading the individual stories where topic 80 is prominent confirms this idea. Where the bush tradition as we know it stands in opposition to domesticity—focusing on largely solitary bushmen with nonetheless intense male friendships or mateship—these stories are more likely to feature families in the bush and to be driven by emotional bonds between men, women and children. The illustration in Figure 5 is indicative of this feature of the fiction, in showing the dangers of the bush—here, as in many of these stories, a bushfire—being confronted by a man and woman together.
The timing of these stories also differs from established understandings of the bush tradition. Where that tradition is identified with the 1890s, these (computationally) characteristic Australian stories are present throughout the second half of the nineteenth century, and indeed, are slightly more common earlier in that period than later. But in terms of existing scholarship on the bush tradition, the most surprising aspect of these stories is their consistent and prominent depiction of Aboriginal characters.

A central argument in Australian postcolonial literary studies is that, beginning in the nineteenth century, fiction replicated the legal lie of terra nullius by not depicting Australia’s original inhabitants (see for example Gelder 2007; Trigg 2005). Yet these bush stories feature multiple Aboriginal characters. Most unambiguously affirm the colonial mission, either by depicting Aboriginal characters in harmonious and friendly relations with colonists, or by presenting them as savage or childish. But others describe complex, multilayered interactions between colonists and Aboriginal people that lead white characters in the stories to question the colonial enterprise. Some depict such radically juxtaposed scenes of violence and purported friendship between colonial and Aboriginal characters that they must surely have prompted the same thoughts in at least some of their white readers. These representations of Aboriginal characters are still profoundly racist. But they demonstrate that the conception of Aboriginal people as excluded from Australian fiction is an effect of subsequent publishing, editorial and critical practices, not of the type of fiction written and read by Australians throughout the nineteenth century.

These four findings demonstrate the capacity of new digital collections and methods to enable new insights into the history of publishing. In this case, they provide the basis for recognising nineteenth-century Australian newspapers—particularly provincial ones—as incredibly active and prolific fiction publishers. These periodicals were connected to global systems—cultural and economic—publishing fiction from around the world and in the case of metropolitan newspapers, actively sourcing it from a range of British and American syndication agencies. But such publishing was also distinctively local, adaptive and independent. Australian newspapers featured an extensive amount of Australian fiction—and editors often foregrounded its presence. These newspapers were part of economic and cultural networks that arose in the colonies to source fiction internationally as well as to provide local opportunities for local authors. And these newspapers published stories that were distinctively Australian: of the bush, but offering a different bush tradition, not restricted to depicting white men.

**Data publishing in publishing studies**

I began this chapter by discussing how digital publishing studies requires us to combine existing disciplinary knowledge and expertise with new research practices relating to digital infrastructure and data provenance and construction. To end, I want to consider another—arguably more fundamental—change we need to undertake when employing digital technologies for publishing studies: data publication. My overarching claim in this final section is that grounding publishing studies on large-scale data requires that we rethink foundational issues of argumentation, sustainability and access.

These issues are fundamentally bound up together. But prior to the digital age, publishing scholars—indeed, humanities scholars in general—have typically made their arguments with reference to evidence held by cultural institutions and/or created by publishers. Thus, we have been able to delegate all but the construction of our argument to others without having to give much thought to sustainability and accessibility. When arguments are based on data, this approach is no longer sufficient. For my project, for instance, I cannot point to Trove and say there is the evidence, because the historical moment in which Carol and I interrogated Trove, and the transformations we subsequently wrought on that evidence, are constitutive of the arguments I am able to offer. Constructing literary data also involves so much labour—and often, so much public funding—and if done well, adds so much value, that we abdicate our scholarly responsibility to the collaborative accumulation of knowledge if we fail to publish, and thus to make available, sustainable and accessible, the basis of our arguments.
Data publishing, then, cannot be an optional extra. But how we perform this activity—and how we involve publishers, cultural institutions and the general public in the process—is something that we, as a field, need to give a lot more thought to moving forward. For the sustainability of my newspaper fiction project, I have lodged all the data I explore in my forthcoming book with the publishers, the University of Michigan Press. This data appears under a creative commons license, alongside an open-access version of the monograph (Bode 2018). But not many humanities scholars can use raw data. The To be continued database attempts to solve this issue of accessibility. It publishes all of the data we collected—not just the titles I investigate in the book—with an interface that enables users to explore fiction in nineteenth- and early twentieth-century Australian newspapers by title, author, gender, nationality, newspaper, and date, and by searching full text records. To be continued is also linked to Trove, so that users can view stories in situ, on the digitised newspaper page. Data can also be exported, with users able to choose whether to download the entire database, or a subset, of bibliographical metadata and full-text records.

With these two forms of publication I feel that—in collaboration with my publisher and my institution—I have fulfilled the requirements of data sustainability and accessibility, at least into the medium term. But there is significant potential to approach data publishing in ways that not only fulfil basic requirements but extend the possibilities of what publishing in publishing studies might offer or enable. I have sought to explore these possibilities in three ways. Most basically, as well as allowing me to publish the basis of my arguments, the To be continued database offers a resource for future arguments. Perhaps a literary scholar is interested in Russian author Ivan Turgenev. She can use the database to explore instances where his fiction was published in the Australian colonies and how it was presented to audiences via the paratexts. To extend the example, she might investigate what other Russian authors are present in the database, and on this basis, explore whether, and if so how, the framework of Russian literature was mobilised in the Australian colonies. A periodical scholar might examine whether the publication of Russian fiction was limited to particular newspapers or widespread, and the social or political implication of this pattern; a textual scholar could investigate the translations used and the insights these offer into the sources of such fiction.

Or to give another example, this time for Australian literary studies, among many other new titles, the To be continued database contains two previously unknown iterations of the city mysteries genre. Begun with Eugene Sue’s “The Mysteries of Paris” in 1842—which was also serialised in Australian newspaper (the Geelong Advertiser and Squatters’ Advocate) four years later [#440]—this genre encompasses adaptations throughout the nineteenth century for multiple cities including Amsterdam, Berlin, Boston, Lyon and more. A “Mysteries of Melbourne Life” from 1872 was known prior to our analysis of Trove’s digitised newspapers; but the two Australian city mysteries thereby discovered predate this publication: by two years in the case of another Melbourne mystery [#14468], and by more than two decades, for the “The Mysteries of Sydney” [#6498] published in Bell’s Life in 1850. These two stories, alone, could represent the basis for a rich research project. As with so much of the colonial fiction discovered in nineteenth-century Australian newspapers, these city mysteries—including the by-line of the earliest, as “Not by the author of ‘The Mysteries of Paris’ or ‘The Mysteries of London’ but by One of Ourselves”—connect to and emerge out of a transnational market, while also emphasising local authorship and distinctively colonial forms of expression.

As already noted, there are over 21,000 novels, novellas and short stories in the To be continued database, and this chapter—like my book—only explores trends in a little over 9,200 of them. For the nineteenth century, this leaves almost 7,500 short stories that, in the vast majority of cases, no one has considered in the Australian context, if at all. As for the 4,500 short and extended stories identified in early twentieth century, to my knowledge no previous research recognises that fiction continued to feature in newspapers at this time, let alone that its publication was so extensive. This collection of fiction, in other words, has the potential to support many more new discoveries in and understandings of the history of fiction publishing in Australia and the transnational circulation of fiction in the nineteenth and early twentieth centuries. To enable this future research, the database is editable, meaning users can improve the information in existing records as well as add new instalments or entirely new stories from Trove’s digitised newspapers.
The second way I have sought to extend what is possible for data publishing in publishing studies is by working with collaborators from Trove, especially Julia Hickie and Victoria Riddell, to capture these records. Trove has now harvested all the stories in the To be continued database, adding over 21,000 records to the NLA’s online catalogue. Any stories that are subsequently added to the database will also be represented in this way. As far as I am aware, this is the first digital humanities project, internationally, to close what we might call the humanities data circle, demonstrating how a major national digital collection can both provide the basis for humanities research and draw on the results of that research to enrich its records. This approach is also another step in the sustainability of research data, in that Trove and the NLA are likely to long outlast the To be continued database.

I have been talking about sustainability and access for researchers. But of course, thousands of members of the public use Trove. Many spend countless hours, in particular, correcting errors in the OCR generated newspaper text. The third and final way I have sought to extend the possibilities of data publishing for publishing studies is by enabling any member of the public to use To be continued and Trove to read and correct fiction, where necessary to identify and add missing instalments, and to discover and add new stories to the database. These stories can be exported from the database, and published in an open-access text repository such as Project Gutenberg; and I am also working with digital artist and developer Geoff Hinchcliffe to create a publication platform to facilitate and enhance this process. There is so much—particularly Australian—fiction in To be continued that has never been published outside of the newspaper pages that members of the public can create their own first edition of a forgotten Australian work. As well as improving the quality of records and text in Trove and To be continued, the crowd-sourcing features of the database present considerable opportunities for making available and curating the works of out of print or marginalised subjects. Alternatively, newly discovered works could be published by a print publisher, as I have done with a collection of previously forgotten works by Catherine Martin (Bode 2017). This is the first in a planned series of newly discovered Australian fiction, with Obiter Press, and possible future collections might include previously forgotten mining fiction, Christmas stories, rural romances or cricket writing (all categories are fulsomely represented in the database).

Clearly, publishing data represents a lot of additional—and different—work to that which we are accustomed to in publishing studies. And it requires that we take responsibility for demonstrating not only the foundations of our arguments, but for making those foundations sustainable and accessible to others: tasks that, as researchers, we have previously been able to relegate to publishers and cultural institutions. However, digital publishing studies can only progress—responsibly and effectively—if we take these steps. And as I hope I have shown, as well as fulfilling a basic requirement, data publication has the potential to forge new connections between publishers, publishing scholars, researchers in a range of academic disciplines and in cultural institutions as well as members of the public. In these ways and others, digital technologies open up new possibilities for publishing studies, including new forms of exploration and understanding, and new communities with whom to share this process of discovery.

References


Editor. 1895. “Marjory’s Mistake’: Another Original Story for the ‘Miner’,” Barrier Miner, April 22, 1895, 3, nla.gov.au/nla.news-article44167084
1 I mention archaeology as a field that seems to generate particularly newsworthy examples of this transformative capacity of digital technologies. Recently reported examples include the Birka Viking, where DNA analysis revealed that a warrior long assumed to be male was female (Nutt 2017), and a crowd-sourcing project in satellite archaeology, where members of the public help to discover previously unknown ancient sites based on satellite imagery (Stinson 2017).

2 The majority—98 percent—of the fiction in this category is extended by serialisation. The remaining extended titles are categorised as such because they comprise of 10,000 words or more in a single (often a Christmas) issue. Some of these are novella-length (10,000 to 25,000 words). But others are full novels of 60,000 words or more.

3 These figures describe the contents of To be continued in November 2017. Since its public launch in March 2018 members of the public have added hundreds of new stories to the database. I discuss this crowdsourcing component of the project at the end of the chapter.

4 Colonial newspaper editors frequently noted the importance of fiction to their readers. For instance, the editor of the Barrier Miner wrote that: “The success which has attended the publication of ‘The Last Signal’ has been so pronounced that the proprietors of the BARRIER MINER have decided to continue, for a time at least, the publication of serial stories; and have just completed arrangements wish [sic] Miss Adeline Sergeant for the publication of her latest brilliant dramatic serial, ‘Marjory’s Mistake'” (Editor 1895).

5 I cite fiction in the To be continued database using the unique record number. Full bibliographical details and available text files of these stories can be accessed by entering this number into the “Trove ID” field in the search interface for the database.