Retrieving a World of Fiction: Building an Index—and an Archive—of Serialized Novels in Australian Newspapers, 1850–1914

Katherine Bode and Carol Hetherington

Two and a half decades ago in this journal Elizabeth Morrison made an impassioned and persuasive case for creating an index to serial fiction in Australian (or Australasian) newspapers.¹ Such an index, she argued, would reveal much about the connections between British, Australian, American and New Zealand literary cultures, and specifically, the influence of these other national literary cultures on Australia’s. Indexes of fiction in specific Australian newspapers and magazines had been created prior to Morrison’s article, as she acknowledged, and others have been published since, all making important contributions to our understanding of literary and print culture.² While this large number of projects—over more than four decades—indicates the desirability of Morrison’s agenda, their history and current state foregrounds what have been major obstacles to achieving this aim. The most obvious of these—demonstrated by the two methods Morrison employed to sketch out the index’s parameters—is the formidable scale of the task. To understand its breadth, Morrison performed a “cross-sectional check” to explore which of Victoria’s one hundred or so newspapers, “issued on or about 31 August 1889, contained instalments of novels.” This method uncovered twenty-eight separate novels—some published multiple times—as well as a pattern of independent publication in metropolitan dailies and weeklies, and syndicated publication in suburban and country newspapers. The second method, to explore the index’s depth, involved “a diachronic study of serials in the Age from April 1872 (when it began to serialise fiction) until the end of the century,” and

² Previous work in this vein includes a range of projects from the 1970s: at the University of Sydney to index literary material in the Australasian and the Sydney Mail; Elizabeth Webby’s study of the Australian Journal; and Laurie Hergenhan’s index of eleven periodicals and newspapers published between 1880 and 1900. In the 1990s Chris Tiffin and the Fryer Memorial Library at the University of Queensland collaborated on an index to the Queenslander from 1866 to 1900; in 2001, Toni Johnson-Woods published Index to Serials in Australian Periodicals and Newspapers: Nineteenth Century (Canberra: Mulini Press, 2001) covering fifteen magazines and weekly newspapers; and AUSTLIT (1988–2001) and its successor AustLit: The Australian Literature Resource (www.austlit.edu.au) have, through various Consortia and individual grants, indexed a large number of national and regional titles, including the Bulletin, the Australian Town and Country Journal, the Melbourne Leader, and numerous Queensland newspapers such as the Cairns Post, the Townsville Daily Bulletin and the Mackay Morning Bulletin, within its “Writing the Tropical North” subset.
identified sixty novels across this twenty-eight year period, mostly by English or Scottish authors, with a few American and Australian titles.\(^3\)

Morrison’s exploratory work showcases the richness of the field at the same time as it highlights the sheer scale of work involved in manually cataloguing fiction in Australian newspapers. Accordingly, despite covering substantial ground, existing indexes leave much more ground uncovered. Their usefulness is further limited by their separateness—which makes it impossible to achieve the large-scale survey Morrison envisaged—as well as restrictions of accessibility and scope. In particular, many of these projects are only fully available as card indexes or microfiche, or are out of print.\(^4\) Where the findings have been digitised, for the most part only information on Australian authors and titles has been retained,\(^5\) an approach that counteracts the “transnational turn” in Australian literary studies since the 1990s.\(^6\) As the history of research in this field demonstrates, conducted via manual methods, Morrison’s case for a comprehensive index to serial fiction in newspapers—while clear and compelling—is also naively optimistic. Even given her admittedly “sketchy” outline of the practicalities, and recognition that the information would need to be held in a “computer database,” a manual search of hard copy or microfilm is unfeasible because of the formidable size of the archive. It would seem, in this case, that bibliographic imagination significantly surpassed what was achievable with the bibliographic methods and resources available.

This situation has now altered profoundly, with new digital resources and methods not only bringing this bibliographic imaginary into reach but enabling us to surpass it in scale and ambition. Specifically, the National Library of Australia’s Trove newspaper digitisation project, combined with automatic

---

\(^3\) Morrison, “Retrieving,” 29.

\(^4\) The findings of the University of Sydney projects, including Webby’s study of the *Australian Journal*, are fully available only as an in-house card index in the English Department; The Hergenhan Index can only be accessed as a card index at the Fryer library; Tiffin’s original computer-based index is now inaccessible; Johnson-Woods’s index is available in print form only (at 14 Australian libraries and currently unavailable second-hand). Further information about the Hergenhan and Tiffin projects can be obtained from the Fryer Library.

\(^5\) Australian literary material in the *Sydney Mail* and the *Australian Journal* was incorporated into *AustLit*; but literary material by overseas writers was omitted. The same is true of the *Queenslander* index: only Australian material was entered into *AustLit*. With some few exceptions, the indexing of Johnson-Woods’s work conducted within *AustLit* excluded non-Australian material. *AustLit*’s online format gives it the capacity to expand and correct limitations of past coverage; a change to its inclusion policy in 2013 means that work by ‘international’ authors will now be indexed, however the nature and extent of its retrospective indexing in this area is not clear. The “Colonial Newspapers and Magazines Project” is contributing to this expanded, international approach through detailed indexing of all of the literary material in a wide range of periodicals in four years: 1836, 1868, 1888 and 1900 (see http://hass.unsw.adfa.edu.au/ASEC/current_projects/colonial-news-and-magazines/index.html).

methods for searching and extracting bibliographic and textual data, enable us to create both a bibliographic index and a textual archive of serial fiction in Australian newspapers. In this article we begin by outlining our methodology, which we believe has significant potential for bibliographical research in general. While overcoming limitations of previous projects relating to physical access to resources and restrictions in the communication of what is found, this new, digital approach presents its own challenges. Instead of how to confront the scale of the task of indexing serial fiction in Australian newspapers—the issue Morrison wrestled with—our questions are more specifically bibliographical as well as more broadly epistemological: what are the bibliographical techniques and challenges of automatically indexing serial fiction in digitised Australian newspapers; and what are we actually representing and knowing when we create such an index? Ultimately, the extensive form of bibliography and view of print culture that these methods enable requires both the adaptation of traditional bibliographic techniques to a new digital environment, as well as new ways of understanding traditional bibliographic entities, including the work, the author, and the archive.

Methodology

The Australian Newspaper Digitisation Program aims “to make freely available online through Trove, as many Australian newspapers published prior to 1955”—and hence, out of copyright—“as possible.” At the time of writing (June 2014) Trove provides access to over 13.5 million pages from over 650 Australian newspapers, with digitised records added daily.7 Trove’s online newspaper collection has been strongly embraced by both the public and the academic community. Marie-Louise Ayres notes that, “By July 2013, use of Trove dwarfed use of the [National Library of Australia’s] other online services, the Trove Newspapers zone dominated use of Trove, and Australians had corrected more than 100 million lines of text—the equivalent of 270 standard work years of crowd-sourced effort.”8 Trove Newspapers has profoundly affected academic research. Julia Robinson describes how the digitised records have “revolutionised” the way lexicographers at the Australian National Dictionary Centre “conduct research into Australian English,” permitting “fast, targeted searches … for a word by date, newspaper, state, text category or combinations thereof, providing an invaluable means of finding evidence.”9 In terms of historical research more broadly, Richard Aitken quips: “Forget Pharaonic epochs, Chinese dynasties, or the birth of Christ, Australian historians now refer to research Before Trove

Examples of how Trove’s digitised newspapers are transforming historical research can be found in a range of areas, including music, literary studies and political history.

Our method aims to revolutionise bibliographical research not in the manner of existing historical research with Trove—that is, searching for specific titles, authors or events, although such an approach is valuable, and part of our strategy—but by using words or phrases that commonly occur within, or are used to frame, serial fiction in Australian newspapers to identify titles. The first term we are using is “chapter,” which has proven very effective in optimising the relevance of search results for serial fiction. In particular, “chapter” often occurs multiple times in the text designated by Trove as an “article” (because a single instalment frequently contains a number of chapters) while also appearing in the “article” title (which includes the first four lines, and is the only part of the digitised text manually checked and transcribed, thus reducing the effects of Optical Character Recognition, or OCR, errors on search results). We will use other search terms as the project continues, including “our serial story,” “our storyteller” and “the novelist.” Each will have its own benefits and drawbacks, and our ultimate intention is to employ a range of terms until the returned results demonstrate a high level of repetition with what is already indexed.

The search term “chapter” returned more than 800,000 results, certainly too many to export manually. For export we use an Application Programming Interface (API) created by Tim Sherratt, now manager of Trove, and accessible through the Trove website. The search URL is entered into the API, which then scrapes the bibliographic metadata and full text records of the results and exports them as csv and text files. The API enabled us to export the following metadata: “article@id” (a unique identifier for each article in Trove), “heading” (the first four lines of the article text), “title” (the name of the newspaper), “title@id” (the newspaper ID, designated by Trove), “edition” (the newspaper’s location and duration), “date” (in the form of day, date, month and year), “page” (page reference), “identifier” (the article’s URL), “trovePageUrl” (the URL for the page where the article is located), and “correctionCount” (the number of corrections made to the article). We also exported the article text in files identified by the same number string as the Article ID.

Due to the usefulness of “chapter” in optimising the relevance ranking for serial fiction, we found that the first 30 or so sets of 5000 results were almost exclusively fiction, with the share of other records to fiction increasing over the next 20 sets of 5000 results. At this point, we deemed the share of non-relevant to relevant materials too high to warrant further investigation of the results. Other results that arose from the “chapter” search included reports of meetings of a chapter of a lodge or religious association, accounts of a chapter in the life of a town or person, or even public documents such as deeds of grant and regulations organised in chapter divisions. It is usually clear from the title of the article whether it is fiction or not, and when this was not the case, we used the article URL field to check the digitised record. The results contained multiple duplicate records, which we removed automatically, based on the article ID number. Any other records that did not qualify, in our judgement, as fiction or literary non-fiction, we also deleted.

While the metadata exported automatically is substantial, it is insufficient for our indexing purposes. The next step was to add fields by manually checking records. Although time-consuming, this process is far quicker than would be the case without the prior automatic extraction of data, particularly as it is more often than not possible to generalise the outcome of a particular manual check to multiple records with the same title except for small changes relating to chapter number and title. We have added a large number of bibliographic fields, including two for title, and two for author name, the necessity of which we discuss in the final section. For records prior to 1899, the period we have indexed

---

13 Some of these bibliographic fields are based on how the publication event is framed in the newspaper and include: “Author of” (for other titles by that author listed in the newspaper); “Trove Source” (for the source of titles identified in the newspapers, such as “Republished from the London Journal”); and “Chapter Number” and “Chapter Title.” Other bibliographical fields are based on subsequent research including: “Gender” and “Nationality” (of the author); “Other Name/s” (used by the author); “Also Published In” (indicating other periodicals we have identified as publishing the same work); and “Additional Information” (on any aspect of the title or author, and its source). A final field identifies titles completed in one or two issues. Like Morrison we focus on “serial (rather than one-off) publication” because it represents a phenomenon “which is sui generis, which has its own characteristics and effects—seen in the agencies and contractual arrangements involved, in the demands, over time, on the form of composition and regularity of submission, in the presumed or documented relationships between text and reader, and so on” (Morrison, “Retrieving,” 28). Yet in some instances, the same work is published in a single issue in one newspaper and across multiple issues in another, and this bibliographic field allows us to retain those one-off publications for subsequent comparison.

14 These title fields are “Trove Title”—the title of the novel as listed in the newspaper—and “Common Title”—the title the work is commonly accorded. If we are unable to identify the title associated with the first published book version of the work—that is, the title traditionally given primacy in bibliography—we use the earliest title that appears in Trove.

15 These author fields are “Trove Name”—the name the author is given in the newspaper, including where it is anonymous, pseudonymous, or unattributed—and “Common Name”—the name the author is commonly accorded.
thus far, the process described above yielded 58,717 unique records, representing 6,269 titles, of which 4,076 are unique (a number are republished in different newspapers). We have been able to identify 1,693 individual authors of these titles, many publishing anonymously; 1,301 titles were by authors we have not yet been able to identify. The number of results and the fact that this is only the first of many searches, indicates the success of this method and the scale of fiction serialised in Australian newspapers.

At the same time, our records are characterised by multiple missing instalments. To explore whether these omissions were caused by problems in our methodology, in the Trove records, or both, we revisited the results for the period prior to 1880, going through each digitised newspaper where we had found serial fiction, manually searching for missing instalments. This process revealed three main causes of omissions: the original publication event (misnumbering, typesetting errors, missing chapters); the state of the archive prior to digitisation (missing issues, illegible and damaged pages); and the digitisation process (OCR errors, zoning issues, some additional omissions). Manually checking the digitised records indicates that our method has the potential to identify the period when a title was published, but cannot always access it completely or know its exact span. Of course, searching for missing instalments of titles we have identified provides no information about the titles we have not found, and thus offers no overall assessment of the reliability of our method. Once the results of our successive searches are showing a high degree of repeatability we will be able to gain a more general perspective on reliability by randomly sampling newspapers to discover if there is serial fiction we have missed, and why.

Bibliography

Indexing serial fiction from digitised Australian newspapers brings many of the same challenges faced by the manual indexer working with hard copy or microfilm sources, relating to the complexity, volatility and ephemerality of periodical publication. While the digital age has produced new resources, tools and techniques for bibliographic research—which enhance and expedite the work and make large-scale bibliographic research feasible—they do not solve every problem, and introduce some of their own particular challenges.

Nineteenth-century literary publishing presents a range of bibliographic difficulties, and this is particularly true with respect to popular serial fiction. Accurate bibliographic representation is difficult enough in the area of book publishing, given the unscrupulous practices and flagrant piracy of many publishers, in both America and Britain, in the unregulated book trade prior to 1891 (when the Chace Act brought the United States into accord with the 1886 Berne International Book Copyright Agreement). Even after this agreement, serial fiction publishing remained unregulated and chaotic. Works were frequently
published anonymously or pseudonymously or with initials or signatures only, and were often reprinted with slightly different titles and attributions. Because much of this fiction was never published in separate book format there may be no possibility of anchoring a text to a particular edition. In his study of syndicated newspaper fiction in the United States, Charles Johanningsmeier describes how “Local newspaper editors could use large, misleading headlines or titles for the fiction, have outrageous illustrations to accompany them, or place them beside whatever material they chose.”16 Readership, expediency and profit were more important considerations for newspaper proprietors and editors than accuracy and literary integrity, resulting in shoddy typesetting, missing episodes and extravagant claims about the originality and exclusivity of the fiction.

The serial fiction in Australian newspapers by non-Australian authors covers a range of genres—canonical works, sensation fiction, dime novels and Penny Dreadfuls—and was reprinted from a variety of sources—either directly or through syndication services, from published books, fiction magazines, story papers and literary journals, or from other newspapers. Sometimes the source was acknowledged, sometimes not; sometimes publication was with the author’s or publisher’s permission, sometimes without. As it moved across national borders, fiction was often plagiarised and rewritten, re-badged and localised.17 The Trove archive, then, presents a body of fictional material that is frequently unidentified or misrepresented and modified in unacknowledged ways, exemplifying what Pamela Gilbert, in her introduction to *A Companion to Sensation Fiction*, describes as the “non-linear, mutating, increasingly rhizomatic structure of Victorian literary, publishing and consumer culture.”18

The fundamental bibliographical task of identifying works and authors in this archive is significantly assisted by new digital techniques, tools and resources, which greatly facilitate the indexing process. In particular, the ‘digital deluge’ provides easy—often free—access to a wide range of resources including: library and bookseller catalogues; digitised novels, newspapers and scholarly journals; digital versions of established reference sources; institutional digital repositories; individual websites from para-academic researchers and family historians; even eBay. The following example demonstrates just how much digital resources assist in bibliographic research. In 1894 a serial appeared in the *Singleton Argus* with the title “Parted Lives: A Tale of Love and Jealousy” and no authorial attribution.

17 For example, American author “Old Sleuth’s” novel *The American Detective in Russia*; or “Piping a Conspiracy” is serialised in several Australian newspapers as “Barnes, the Australian Detective.”
A quick title search of Worldcat yields no results, but a search using keywords from the text reveals that it is substantially the same text as the serial “After Many Years,” running during 1898, again without attribution, in the Goulburn Evening Penny Post, and “Estranged; or, The Lost Heiress of the Champneys,” published in nine Australian newspapers between 1912 and 1921, again with no author named but with the signature “By the author of ‘The Hampton Mystery’ and ‘Sir Peter Eldon.’” Searching for these titles in library catalogues proves fruitless but a return to Trove finds “The Hampton Mystery” attributed to the prolific American writer Mrs Harriet Lewis. Thus, digital resources—and keyword searching in particular—very quickly yields a series of relevant resources that could only be discoverable serendipitously by manual searching and, if found, would take many years to uncover.

But which, if any, of these titles is correct? Or more to the point, which shall we use to collect the several serials together and provide adequate bibliographic information for a researcher looking for the work in a wider context? Although we are not concerned with establishing an authoritative text in the sense that this term is used in close reading, we still try to find the title associated with the most authorially-controlled and persistent version of the text: that is the first published book version traditionally given primacy by bibliographers. A keyword and phrase search in Google takes us to the New Zealand digitised newspaper archive Papers Past, which contains a substantially similar version of the text, running as “Saltair Manor” in the Taranaki Herald (and other New Zealand papers) in 1873. Lewis is not named here as the author, but there is a signature: “By the author of ‘Honor Glint’, ‘Jasper Lowder’ and ‘The Haunted Husband,’” the last published New York: R. Bonner’s Sons, 1893. A later New Zealand publication running in 1885–1886 has the title “The Sundered Hearts; or, The War of the Household” (here Lewis is named), and even later still, in 1912–1913, a substantially similar version of the text appears in New Zealand, in the North Otago Times, with the title “Estranged; or, The Lost Heiress of the Champneys” (perhaps from the same syndicated source as the Australian publications of the same date). The Google search also points to a serial running with this title in the Courier (Middlesex) in 1887–1888 (found in the commercial newspaper database newspaperarchive.com).

Returning with this information to Lewis’s entry in WorldCat uncovers a novel published with the title Sundered Hearts (New York: R. Bonner’s Sons, c. 1889)—and for the purists, we can say that the title page of this first edition has been sighted (and the image captured as a screen shot) courtesy of a vendor on eBay. With this title, more information can be uncovered: including a version of the text, published from 11 June to 1 October 1870 in the now digitised index to

---

19WorldCat is a union catalog that itemizes the collections of 72,000 libraries in 170 countries and territories which participate in the Online Computer Library Center (OCLC) global cooperative” (Wikipedia).
Retrieving a World of Fiction

the London Journal (compiled by Andrew King, part of the Victorian Research Guides series)\textsuperscript{20} with the title “Disunited; or, Love and Jealousy.” A footnote in this index informs us that the title was also published in the New York Ledger 21 May to 10 September 1870 as “Sundered Hearts,” almost certainly the text’s first appearance. An added bonus from the same search is finding an image of the New York Ledger issue with our serial on the cover, digitised as part of the Dime Novels and Penny Dreadfuls site at Stanford University.\textsuperscript{21} From a number of sites we can confirm the gender and nationality of the author. Excellent though King’s index is, and while we use it routinely, we would not have made the connection between “Divided Lives,” “Sundered Hearts” and the three serials in Trove without the aid of powerful searching tools and digitised material. As it is, we have been able to discover substantial bibliographical information about the work in a relatively short time, using resources physically located in three continents, without leaving the computer.

Although this process may seem laborious, all of this information was gathered with a couple of quick searches of the open web, and has probably taken longer to explain than to actually complete. While the above would be enough to make a respectable entry in our bibliography, if we sought further information about the Lewis novel we could pursue other avenues of research. Thus, a keyword search of ProQuest’s British Periodicals Online database identifies three other publication events: in the London Reader, as “Lady Barbara” in 1870, simultaneously with its serialisation in the London Journal; and reprinted in the London Journal in 1894 as “Love and Jealousy,” “By the author of ‘The Heir of Everton Court.’”

Not all our efforts are as successful as the one described above. Take another example: the serial “Josephine’s Divorce,” published without attribution in the Brisbane Courier and the Queenslander in 1867, and in the Launceston Examiner in 1897 with the title “Against the World; or, A Woman’s Struggle,” “By the author of ‘Naomi’, etc.” Searches for these titles have to date proved fruitless. The serial may be by Annie Webb, whose 1841 novel Naomi appears to be well known, but no confirmation is possible at this stage. A online keyword and phrase search has found the text serialised in the Belfast Morning News in 1868 as “Self-Divorced; or, A Woman’s Stratagem,” and in the Lincolnshire Echo in 1893 and the Derbyshire Times and Chesterfield Herald in 1895 as “A Woman’s Will; or, Energy Rewarded.” But in all of these cases there is no indication of authorship. We can probably assume, because of the publication date, that it is not an Australian work (although British and even American publication does not automatically confirm this). Apart from this we cannot hazard a guess at the author’s nationality and there is no indication of gender.


The research process described above is cumulative, sometimes unpredictable and often serendipitous; the methodology is not new, it is traditional scholarly detective work. But it is made easier, and possible on a larger scale, by new digital resources and techniques, which provide new ways of accessing, and thus of understanding, the complex networks implicated in newspaper publication in the nineteenth and early twentieth centuries. Google Books has been justifiably criticised for remediating the cultural archive in ways inimical to humanities scholarship, because ignorant or uncaring of correct bibliographical and book historical protocols. Yet we can see here how the extensiveness of the cultural archive we have access to via another area of Google—its search interface—provides a major boon to bibliographic research.

There are, of course, frustrations and limitations. Searching the open web is fast but not always reliable or conclusive. Searches can be unsuccessful due to poorly scanned material and bad OCR, which can make finding the right phrases and keywords a matter of chance rather than design. This is particularly the case with non-academic online information; however, financial pressure on libraries has also diminished the authority of their catalogues. There is still no substitute for physical sighting of an artefact—something not always possible within our current project, but not impossible in the future. The priorities of the search engines themselves are another frustration: results are becoming increasingly cluttered with material from social media and commercial sites, rendering selection and appraisal of results difficult as scholarly or not-for-profit sites are relegated to ever lower positions on the results screen. Many valuable electronic sources, such as The Wellesley Index, ProQuest’s British Periodicals Online and the British Library’s British Newspapers archive exist behind high-cost subscription barriers. In addition, these and other databases must be accessed and searched individually.

The plethora of material the online environment returns has challenged us to revise our bibliographic schema. Where initially a “verified in” column seemed


23 The majority of searchable full text databases, including Trove, present material both as a scanned image and as OCR text, and while the scanned image is frequently very readable, the OCR text can be wildly inaccurate. Search engines search the OCR text. A particular example: the novel “Dan’s Treasure; or, Labour and Love,” by Leigh Tempest, was reviewed in The Spectator but appears there as “Vat’s Treasure,” so the title was of no use as a search term. <http://archive.spectator.co.uk/article/20th-april-1867/22/vats-treasure-or-labour-and-love-by-leigh-tempest>.

appropriate for recording a reliable, recognised source for all our attributions, in many cases the authors and titles we are discovering are not covered in these sources. Instead of relying on authoritative sources, it has been necessary to build a justification for assigning metadata using partial information from a wide variety of sources, including newspaper reports, fanzines, family history sites and advertisements. Some of these sources are of questionable accuracy; some may disappear from view in future. The necessity of making judgements based on clues and fragments places the emphasis on recording the basis of our judgements. To this end, we have collected all information sources, either as citations or web captures, so that an independent future assessment of our conclusions will always be possible.

Epistemology

As the discussion above illustrates, in building this index we are faced with a body of—often incomplete—serialisations with numerous alternative titles, under various authorial guises. Now that it is possible to use automatic search and retrieval methods as the basis for creating an index to serial fiction in Australian newspapers, we are confronted by important epistemological questions regarding this task. In particular, what are we representing when we create such an index, and on this basis; and to what extent does this index allows us to know print culture? Addressing these questions foregrounds three epistemological principles that underpin our project: first, there are different aspects of print culture that such an index has the potential to represent, and it is important that discoverable bibliographical facts do not replace bibliographical traces indicative of the contemporary reading experience; second, indexing serial fiction in Australian newspapers emphasises the abstract nature of the work and of the author, and requires a model that represents print culture at both abstract and documentary levels; and third, because we can never access the historical record in its entirety, and because new digital resources and methods complicate such access at the same time as they increase it, this index and full-text archive needs to be understood in terms of provisionality and perpetuity, rather than completeness and totality.

Answering the questions “who wrote it?” and “what is it called?”—to supply the author and title of a work—are at the basis of bibliographic research, and provide the route to further bibliographic facts, such as previous or subsequent publications of a work, the gender and nationality of the author, and so on. Identifying these facts regarding the documentary record can lull us into believing that we are accurately representing the instances of serial fiction

---

25 In the first example, we have a text, or a series of related texts, which has appeared with eight variant titles, only one of which has the same title as an authored, published and copyrighted work. In the second example, the text(s) appear(s) with four different titles and no indication of authorship or separate publication.
we have identified. And in one respect we would be: to return to the agenda Morrison outlined, knowing the author’s nationality, for instance, enables us to ask questions about the transnational origins of fiction in Australian newspapers, and the influences of different national literary cultures on Australia’s. However, the facts about the transmission and publication of print culture are different from—and should not be substituted for—the bibliographical traces of the experience of contemporary newspaper readers.

The frame of reference through which contemporary readers experienced serial fiction in newspapers differs from the frame of reference signified by subsequently discovered bibliographical facts about that print culture. For instance, as noted above, with novels often published anonymously or pseudonymously, or with signatures that emphasise the work over the author (“By the author of …”), knowledge of the author’s name—and in relation to this, their nationality or gender—was not necessarily part of the contemporary experience of reading serial fiction. In other cases, including where the author’s name was not given, nationality or gender was emphasised, bringing these aspects of the work’s origin to the foreground of readers’ experiences. Rather than overwriting these bibliographic markers with discovered bibliographic facts, we retain the information presented in the newspapers alongside the information we subsequently discover.

Another way of expressing the above is to say that we index serial fiction in Australian newspapers at the abstract level of the work and the author, as well as at the documentary level of specific, bibliographic objects. This terminology and framework draw on Paul Eggert’s discussion of the relationship between “work” and “text” in *Biography of a Book*. As Eggert explains, we refer to “the work,” but we experience and interpret a bibliographic object representative of a specific publication event and version. Many actors intervene in and shape each publication event and version, from author and publisher to editor and typesetter, and the reader’s experience and interpretation are affected by the particular text that appears as well as its presentation (in the newspaper, this could include the size of margins and columns, the presence or absence of illustrations and other articles or advertisements adjacent to the serial fiction, as well as any text introducing or positioning the novel).

Because “the work” exists as a series of material documents, there is, Eggert states, “no such thing as ‘the work itself.’” The concept remains “pragmatically

---

26 For instance, fiction serialised in the *Maitland Mercury & Hunter River General Advertiser* was typically introduced with the phrase “From English, American, and other Periodicals.” On the other hand, other publications emphasised the “originality” and Australianness of the fiction. Some pseudonyms are designed to align the unknown author with a specific gender, such as “A Mildura Lady” or “A London Man.”


28 Ibid., 9.
indispensable” in providing us with a name for the thing we describe or experience, but “the work” does not exist as a stable entity. Rather, it is a process that unfolds over time, with “the work” and its particular, documentary manifestations “always in a dialectic: they need each other … to secure their different but linked identities.” The author represents a different case to the work, in that, unlike the work, there is usually—though not always—a person in the world who produces particular documents we come to know by the names of particular works. However, Eggert’s framework can be usefully extended to consider “the author” as another abstract concept. That is to say, we use the author’s name to stabilise the notion of the work and our capacity to interpret it, but each publication event and version involve multiple actors, rendering authorship also a process that unfolds over time rather than the stable product of a specific individual.

Our model links the abstract concepts of work and author with specific documentary manifestations to represent fiction serialised in Australian newspapers in terms of processes rather than stable products. Each bibliographic object is indexed with the title and author name that appears in the newspaper: their “Trove Title” and “Trove Name” (which might by unattributed or “Anonymous”). Bibliographic objects with a substantially similar text are accorded the same “Common Title” and “Common Name,” and thus linked together via the abstract concepts of work and author. These abstract concepts provide the basis for conceptualising relationships and connections within print culture (for instance, works by the same author, or by authors of the same country or gender). Yet linked to their documentary instantiations, the existence of these works as temporal processes is made an inherent part of the analysis.

Constructed in this way, this index and full-text archive allows us to ask questions that distinguish between the contemporaneous experience of print culture and its broader operations: for instance, to explore the proportions of serial fiction from different countries published in Australian newspapers, while also considering the different markers—both within and surrounding the text—that served to make these national origins visible (or not) to readers of the time. Our differentiation of the abstract work from the bibliographical

29 Ibid., 10.
30 Ibid., 15.
31 In some cases, the author is a construct. In addition to pseudonymous identities used by individual authors, our archive includes fiction published under house names that American and British magazine publishers used for fiction written by a number of different authors. Street & Smith’s Nick Carter is a prime example: a fictional character who was also the ‘author’ of the books he appeared in (WorldCat lists several hundred titles “by” Nick Carter written by John R. Coryell as well as dozens of other authors over a long period). Bertha M. Clay is another Street & Smith house name, as is Clementine Montagu(e) in the London Journal (for discussion of this practice see Andrew King, The London Journal, 1845–83: Periodicals, Production and Gender (Aldershot, Hampshire; Burlington, Vermont: Ashgate, 2004), 160–61).
object also enables questions relating to the transformation of works over time: for example, analysis of the multiple titles and author attributions under which particular works were published. As these questions indicate, this index and full-text archive of serial fiction in Australian newspapers presents significant opportunities for exploring this rich and complex area of print culture, much as Morrison envisaged. However, in asking such questions, it is vital to remain aware of the provisionality of this record and of the insights it enables.

All archival research is provisional, in that we never access the full documentary record, whether because it has been misplaced, mislabelled, damaged, lost or destroyed. But digital methods and resources increase the potential for unrealised disjunctions between our perceived and achieved access to the archive. They do this by increasing the proxies or models involved in all archival research. In traditional book historical research, each print cultural document represents countless other print cultural documents (such as the many newspapers of the same title and date, the vast majority of which no longer exist) and is accessed via a particular collection model, such as a library card catalogue. Automatic search and retrieval methods introduce many additional proxies into these processes of representation and access. In the case of our project, the csv and full text data we extract are models for the OCR-rendered digitised newspaper pages collected by Trove, which in turn, are models of the newspapers—or of the microfiche models of those newspapers—held in library archives throughout Australia. Trove’s search interface not only provides a system for accessing these documentary proxies, but mediates such access via multiple other models. For instance, each search result or collection “view” “has its own home page, its own relevance ranking algorithm, and its own facets, influenced by the type of material included in the view.” As the number of proxies or models multiplies, so too does the potential for any changes in these representational systems, or any biases or errors they introduce or perpetuate, to alter how the collection is accessed and represented.

Some changes or problems—such as the digitisation of additional documents, or biases in the search results because certain titles are not digitised—might be relatively easy to identify, and hence, to accommodate. The same cannot be said of the processes that construct, and frame our access to, the archive. In particular, the effects of OCR and search algorithms on search results can be difficult to perceive and rectify. Sampling can reveal specific ways in which these technologies misrepresent or omit aspects of the archive. But one can never be certain that all issues have been dealt with because, without these devices, our

32 When Morrison surveyed the colonial Victorian press, of the almost 400 country newspapers she was able to identify, she discovered only incomplete files for many, and no copies at all for approximately five per cent.
only access to the archive is manual, returning us to difficulties presented by its sheer size. The importance of emphasising provisionality in respect to digital projects is further increased by the rhetoric of objectivity and comprehensiveness that attends such research, where the very scale of information recovered, as well as the supposedly direct access to an underlying system provided by computers, encourage a perception that all has been revealed.

Ultimately, at the same time as they massively increase and enhance our understanding of the serial fiction published in Australian newspapers, digital resources and methods introduce uncertainties into this process of knowing. In *A New Republic of Letters*, McGann reflects that we can never know the meaning of a document or tell the truth about an archive because we can never know the entire system—including the current moment of interpretation—that the document or archive connects to. The “radical uncertainties”34 of the textual condition call, he argues, for a “scholarship of indeterminacy”35 wherein our task is not to know, but continually to thicken our knowledge of, the dynamic network that constitutes print culture in any time or place. This conception of the philological process would seem especially pertinent to the application of digital methods to the digitised newspaper archive. The uncertainties in this process do not replace scholarship, and specifically, its rigor and method. Simply because we can never know the truth of the archive does not mean we should not double-check our search and extraction process, or hunt down the various versions of a story and details about an author, or attend to the differences between documentary and abstract features of print culture. The scholarship of indeterminacy is not the despair of relativity; it represents an argument for rigor and method, because these are the only ways we can progress: by gradually thickening what we know and engaging us in a continual process of reflecting on how it is we know it. In adding detail after detail to our picture of fiction serialised in Australian newspapers, this index and full-text archive, and the digital methods that enable its construction, provide the means by which to understand progressively more about the complexity, richness and cultural importance of this key aspect of print culture.

*Australian National University and the University of Queensland*

---

35 Ibid., 158.