

## A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History

*Advances in computer technologies have made it easier than ever before for historians to access a wealth of sources made available in the digital era. This article investigates one way that historians have engaged with the challenges and opportunities of this 'infinite archive': distant reading. We define distant reading as an umbrella term that embraces many practices, including data mining, aggregation, text analysis, and the visual representations of these practices. This article investigates the utility of distant reading as a research tool via three newspaper case studies concerning Muhammad Ali, women's surfing in Australia, and homophobic language and Australian sport. The research reveals that the usefulness, effectiveness, and success of distant reading is dependent on numerous factors. While valuable in many instances, distant reading is rarely an end in itself and can be most powerful when paired with the traditional historical skills of close reading.*

**Keywords:** digital history; digital humanities; distant reading; homophobia; methodology; Muhammad Ali; newspapers, women's surfing

Conceptualizing history-making in the digital era is complex.<sup>1</sup> At one level, what we might term *Digital History and the Archive*, there is an ongoing, large-scale digitization of material which has provided unprecedented access to an ever-increasing archive, but which also has generated a raft of ideological, institutional and political issues. In this context, historians are encouraged to entertain the idea of working with data, data-driven techniques and visualizations.<sup>2</sup> At another level, what we might call *Digital History as Archive*, historians have access to social media which not only facilitate communication amongst scholarly communities, but can assist historical work by providing material for analysis from Facebook, Instagram, Twitter, YouTube and blogs.<sup>3</sup> Social media also brings with it a host of methodological, ethical and legal issues.<sup>4</sup> At the remaining level, what we might call *Digital History is History*, there are new platforms for history-making.<sup>5</sup> These platforms may be collaborative through community involvement, they may embrace the publish-then-filter model of knowledge creation, they may utilize multimedia, they may only exist online with no equivalent print version, and they may not be

wedded to linear narratives.<sup>6</sup> In these forms, digital history challenges the very nature of traditional historical scholarship.

In this article, we specifically address the first part of our tripartite conceptual tool, *Digital History and the Archive*, discussing concepts such as computer analysis, databases and metadata. Some readers may turn off at this point, as history is considered a qualitative discipline of the humanities rather than any exercise in number crunching – that is the mantra of the sciences. Yes historians have used numbers; occasionally, and sparingly, particularly in specific historical moments epitomized by the Annales School, Cliometrics and the experiment in the 1960s and 1970s with historical computing,<sup>7</sup> and in the case of sport history through the scholarship of Wray Vamplew.<sup>8</sup> But for many, history is a qualitative endeavour characterized by a carefully crafted, written narrative. Indeed narrative is the main game for historians.<sup>9</sup> Understanding narrative as a form of written expression has moved from an accurate assessment of the past to a particular version of events inescapably shaped by the historian.<sup>10</sup> If anything, there is a greater appreciation of the literary qualities of narrative and the role played by historians in the history-making process. Historians are starting to openly invest in understanding form and content, subjectivity and reflexivity, and moral and ethical positions, as they seek to represent the past.<sup>11</sup>

If historians are investing more in understanding the literary qualities of their work, what has prompted reflections about using forms of quantitative analysis to inform history-making? What has driven an essential U-turn in the focus for some historians? The change is the unprecedented volume of data available in the digital era. In 2012, there were approximately 21 billion indexed web pages, Google had scanned 14 million books and indexed a trillion web pages, JSTOR provided access to seven million articles from 1,000 publishers, and Facebook provided five million pieces of content weekly.<sup>12</sup> Major national repositories, like the Library of Congress, the British Library and the National Library of

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

Australia, have diverse and growing digital collections. These resources are referred to as the 'new infinite archive', a term that attempts to acknowledge the burgeoning, potentially limitless, and readily available digital traces of the past.<sup>13</sup> For historians, this is a new world, a new economy of knowledge: it is a world of abundance, not scarcity.

While abundance abounds, the same issues influence digital archives as they do traditional archives. Who decides what records are copied: archivists, academics, government bureaucrats, commercial operators? Is it possible, or even desirable, to copy and store everything? If not, what records will be stored, which will be lost and forgotten? How are these records stored, and by whom? Will it be public institutions – government, libraries and museums – and/or commercial companies? Will the records be publically available, or only accessible for a fee? If access is fee-based, will this entrench the distinction between institutionally funded scholars and amateur historians? Finally, will those records which require a fee, and those that are not available at all, take on a different status based on their scarcity?<sup>14</sup> In this sense, the same argument put forward about traditional archives by Foucault, Derrida and, more recently, Steedman hold true for digital archives: they are sites of power for states, organizations and corporations.<sup>15</sup> Digital archives – not unlike the historical development of traditional archives – are shaped by ideological, institutional, political and financial factors.<sup>16</sup> The challenge for historians, though, remains the same: to determine authenticity, trustworthiness and reliability of the archive, and to convey how this largely interpretative dimension shapes the subsequent historical narrative.

But there are important differences. A historian's experience in the digital archives is a world away from directly working in information institutions: waiting for research material to be recalled from the stacks and storage areas, sorting through boxes full of pamphlets, minute books and other documents, or loading copies of newspapers onto microfilm readers to read them with strained eyes,

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

and writing down the discoveries of the day, week or year in personal diaries and journals. This kind of embodied experience of archival research, what Robert Darnton refers to as 'a kind of marinating', has been central to the history-making process.<sup>17</sup> This includes not only the aesthetic and sensorial dimensions of archival work but also the importance of 'original' primary sources that is beautifully captured by Arlette Farge.<sup>18</sup> The digital archival experience is devoid of the 'original' primary sources that Farge rejoiced in, offers very few aesthetic and sensorial dimensions, and engages in a different kind of marinating than Darnton referred to through online searches of documents carried out, most likely, on an office computer.

One of the key issues for historians is the transition from hard-copy remnants of the past to digital copies on a computer screen. Communication theorists have contended for a long while that 'media do not simply convey messages, they affect our very relationship within the world.'<sup>19</sup> Digitized newspaper articles, the focal point of this article, are a prime example. As Bob Nicholson has argued: 'Though a digitised text might look familiar, it is not the same source; we are able to access, read, organise and analyse it in radical new ways'.<sup>20</sup> Traditional research of newspaper articles invariably involves visual scanning of hard-copy pages or microfilm copies looking at whole editions, special sections, headlines and images followed by closer analysis of articles. In digitized archives, this 'top-down' approach is tipped on its head and becomes 'bottom-up' computer-generated searches. Search interfaces, often involving Optical Character Recognition (OCR), locate words and phrases in the infinite archive. The benefits are obvious: 'Billions of individual words – the fundamental building blocks of culture – are now at our finger tips'.<sup>21</sup> Never before have historians had this unique combination of a massive digitized archive and the tools, including search interfaces and computer-generated programs, to access, engage with and evaluate such a vast array of historical material. What we have available in

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

terms of the infinite archive and computational power 'represents a moment of revolution', as Matthew Jockers argues, in the way we study the past.<sup>22</sup>

As much as digitization opens up, even demands, new research approaches, it raises a number of problematic issues for historical work.<sup>23</sup> Search interfaces or computer reading machines, the crucial link between digitized sources and historians, are designed by libraries, archives and companies. Historians rarely have input into their design, often creating less-than-ideal research tools and these often bespoke search interfaces require skills not traditionally found in the training of professional historians. When analyzing newspapers, computer reading machines employ keyword searches to reveal the number of articles with positive results. This process might sound simple but terms can be hard to identify because words may be spelled in different ways, words often have multiple meanings, and the meaning associated with words changes over time. In acknowledgement of the complexities of language and the demands of working with large amounts of data, there are often a range of search options, including Stem, Fulltext, Proximity and Boolean, which help locate appropriate words, combination of words and phrases. Beyond the fact that this requires upskilling for historians, search interfaces for newspapers, like OCR, are far from perfect. Unlike the success of computer reading machines used to analyze eighteenth- and nineteenth-century novels, OCR in historical newspapers is nowhere near as successful as the transition from hard copy to digital forms is affected by the print quality of the original. In recognition of this problem, many information institutions have crowd sourcing projects that manually correct errors in digital copies of newspapers.<sup>24</sup> It is essential to acknowledge that the success and applicability of digitization and computer reading machines differs between fields of inquiry and 'takes place in an economy of loss and gain'.<sup>25</sup>

If we accept that digitization of archival material and search interfaces offer unique advantages, as well as genuine limitations and several unexplored issues, what can be gleaned from other academic

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

fields to help historians make the most of the infinite archive? Literary studies provide some helpful insights. A decade ago, Franco Moretti used quantitative methods to examine 7,000 British novels of the eighteenth and nineteenth centuries.<sup>26</sup> Moretti created and popularized the concept of 'distant reading', which refers to assessing descriptive data (metadata) that is derived, but one step removed, from the textual content.<sup>27</sup> Recently, historians and sport historians have engaged with distant reading as a way of enabling the investigation of large amounts of historical newspapers that are now available through major commercial databases like *19<sup>th</sup> Century U.S. Newspapers*, the *17<sup>th</sup> and 18<sup>th</sup> Century Burney Collections*, and ProQuest's *Historical Newspapers*, and non-commercial projects including *Chronicling America*, *New Zealand's Paper's Past* and *Australia's Trove*.<sup>28</sup> In this article, we engage with some of these newspaper databases and employ distant reading as an umbrella term that embraces many practices including data mining, aggregation, text analysis and the visual representations of these practices in the forms of charts and graphs.

In this article, distant reading of digitized newspapers is used to explore three case studies. The first case study details research into Muhammad Ali, the American boxer and icon of racial politics; the second is an analysis of the emergence of women's surfing on the beaches of Australia at the turn of the twentieth century; and the third is the examination of homophobia in Australian sport from the late 1800s. As much as these examples highlight the infinite archive, we will focus on the exploratory phase of research methods as historians who are experimenting with distant reading. We engaged in this experimentation based on three premises: the necessity to understand the rhetorical dimensions of working with quantitative approaches, the importance of making research assumptions explicit, and opening up data to other historians to refine, rework and contest our findings.<sup>29</sup>

*Case Study: Muhammad Ali*

This first case study explores a distant reading of the boxer Muhammad Ali in the *Los Angeles Sentinel*, a 'black' newspaper based in Los Angeles.<sup>30</sup> Digital copies of the *Sentinel* were accessed via *ProQuest Historical Newspapers*, an online, subscription-based archive that houses a number of North American and international newspapers.<sup>31</sup> Ali's presence in the *Sentinel* was part of a larger distant reading project that studied how twelve North American newspapers wove racial and religious discourses into their writing about the boxer between 1964 and 1975. This section outlines the conceptual, historical and epistemic issues that underpinned arguably the most crucial decision in the project: choosing which searchable words or phrases to use as the foundation for distant reading.

There are a number of ways to conduct a distant reading, ranging from extremely complex, algorithmic analyses of semantic relationships within texts to more simple methods. This particular case study was based upon a relatively simple premise, using ProQuest's inbuilt word-search tool to find particular words or phrases within the *Sentinel* articles that might illuminate racial or religious discourse related to Muhammad Ali. The frequency with which these words appeared within the *Sentinel* was charted over time in order to highlight trends, patterns or themes related to press attitudes toward Ali during the most significant years of his career (1964-1975). But which words would be most likely to illuminate these trends? Selecting search terms which would yield meaningful results was grounded in knowledge of Ali's racial ideologies, his religion, his tempestuous relationship with the press, and his place within broader social discourses throughout the 1960s and 1970s.

It was decided that the search terms which best embodied these considerations were Ali's two names: 'Cassius Clay' and 'Muhammad Ali'. Although these may seem like an obvious choice, their centrality to this project was made by striking a balance between conceptual, historical, pragmatic and

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

technical issues. Firstly, both the names are unique to the topic at hand. A search for 'boxing' and 'race', or 'Islam' and 'Civil Rights', would likely have uncovered articles related to the racial and religious discourses being addressed in this project, but there is no guarantee that these articles would be specifically about Ali. Neither 'Muhammad Ali' nor 'Cassius Clay' were names in common parlance during the 1960s, and there was certainly no other public figure of the same name whose fame even approached that of the man in question. Using these two names in a Fulltext (exact) search ensured that the articles returned were at least specifically related to Muhammad Ali, the boxer. Furthermore, the uniqueness and length of the two names also ensured that OCR software was unlikely to mistake them for other words. This was further aided by the fact that *ProQuest Historical Newspapers* claims an incredibly high rate of OCR accuracy: 99.995%.<sup>32</sup> Beyond these pragmatic and technical advantages, the most compelling reason for selecting the boxer's two names as the building blocks for this project lies in their deep, discursive links to Ali's blackness and his religion.

In March 1964, newly crowned heavyweight-boxing champion Cassius Marcellus Clay Jr. changed his name to Muhammad Ali. The name change symbolized Ali's official conversion to the Nation of Islam, a controversial religious sect renowned for advocating black-nationalism, black autonomy and the separation of black and white Americans. The timing of his name change magnified its symbolism: Ali became a "Black Muslim" in the middle of the Civil Rights Movement's most active years.<sup>33</sup> By aligning himself with the Nation of Islam, Ali placed himself in direct opposition to large sections of the population, both black and white, who were working toward integration. As Mike Marqusee states, 'this was a Black man signalling by his name change, not a desire to ingratiate himself with mainstream America, but a comprehensive rejection of it.'<sup>34</sup> In the Northern states at least, mainstream America generally supported the Civil Rights Movement and this included a great number of newspapers.<sup>35</sup> Journalistic attitudes toward the Civil Rights Movement in the 1960s and 1970s are far too complex and

diverse to be discussed in full here, however, by the time of Ali's name change, many newspapers found Ali's beliefs to be divisive, extremist and potentially damaging to the progress being made by Martin Luther King Jr. and other Civil Rights leaders.<sup>36</sup> As a result, a great number of journalists refused to use the name Muhammad Ali in their columns; instead they called him Cassius Clay as a way of protesting the racial, religious and political ideologies associated with his new name.

The discursive and ideological 'baggage' that Ali's two names carry is the key to their effectiveness as search terms. Distant reading of the *Sentinel* counted the number of articles containing 'Cassius Clay' each month between 1964 and 1975 and compared it to the number of articles containing 'Muhammad Ali' in the same time period. The comparative dimension of this distant reading is vital and is predicated upon the hypothesis that journalists used either 'Cassius Clay' or 'Muhammad Ali' in their writing to make a discursive point about his racial or religious ideologies. Comparing *Sentinel* journalists' usage of the two names allows us to detect trends, patterns or themes that illuminate how attitudes towards Ali, and these ideologies, evolved throughout the 1960s and 1970s.

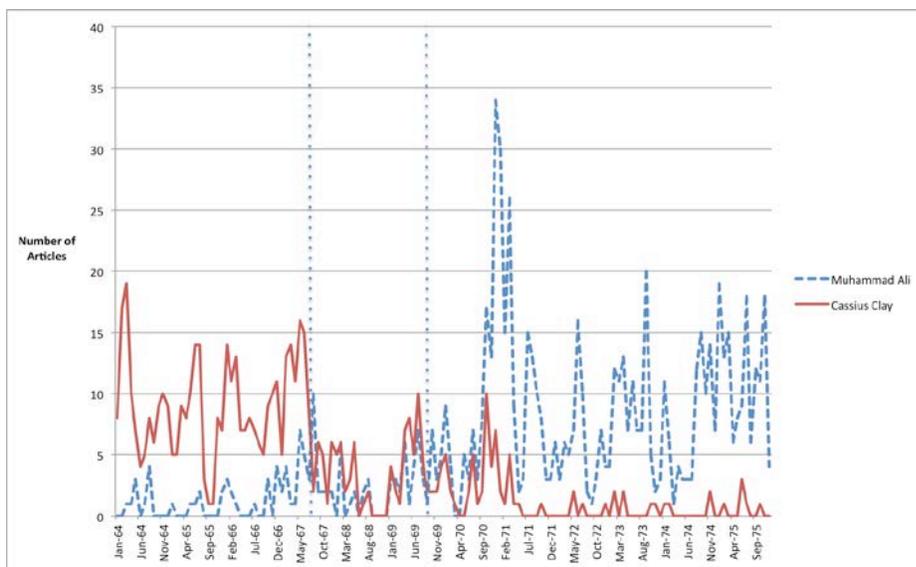


Figure 1: Number of *Los Angeles Sentinel* articles containing 'Muhammad Ali' or 'Cassius Clay' (1964-75)

Figure 1 highlights a number of distinct trends and shifts in the usage of both 'Cassius Clay' and 'Muhammad Ali'. Firstly, it is clear that between 1964 and late 1967, *Sentinel* journalists overwhelmingly referred to Ali as 'Cassius Clay'. Secondly, between late 1967 and late 1969, the two names are used far more interchangeably. Thirdly, after late 1969 (and more specifically by the end of 1970), the name 'Muhammad Ali' is used more often than 'Cassius Clay'. These three periods are demarcated by distinct shifts in the usage of each name (marked with vertical dotted lines) and indicate that the attitudes of *Sentinel* journalists towards Ali shifted decisively throughout the period in question. These trends and shifts generate lines of inquiry about changes in the use of Cassius Clay and Muhammed Ali that are indicative of, and contribute to, discourses about race, religion and politics.

#### *Case Study: Australian Women's Surfing*

The historiography of Australian surfing is virtually silent on women's involvement in the development of surfboard riding. There are some notable exceptions,<sup>37</sup> but academic writing on surfboard riding has focussed on male participation, largely in pursuit of foundational chronologies and myths.<sup>38</sup> One woman who is singled out is Sydney's Isabel Letham for supporting Duke Paoa Kahanamoku's surfboard demonstration in the summer of 1914-15.<sup>39</sup> Beyond Letham's story, very little is known about the popularity of women's surfboard riding in Australia during the twentieth century.

The digitization of Australian newspapers provides the unprecedented opportunity to examine women's surfing and surfing more broadly. As soon as this opportunity was explored, we realized that every dimension of our distant reading was shaped by conscious and deliberate decisions about the digital research process, as well as factors outside our control. Sydney is recognized as the hotbed of early Australian surfing and the search could have legitimately been limited to newspapers from this

geographical area. However, to capture the widest representation of surfboard riding, we searched for instances of surfing from all Australian States and Territories in 900 newspapers via *Trove*. *Trove* enables searches of the entire content of newspapers, including advertising, lists, results, guides and family notices, but we only included articles as we were specifically interested in reports of the activity. Finally, the digital search was restricted by a factor outside our control: Australian copyright laws prevent the digitization of newspapers of Australian newspapers beyond 1954 and thwart the ability to perform any distant reading in the second half of the twentieth century.<sup>40</sup>

We experimented with all forms of searches. In the end, Fulltext (exact) and Proximity searches were excluded because the first does not identify some important word extensions, and as much as the latter reduced the number of false finds, it did not guarantee search relevance and omitted key articles. We opted for a combination of Stem and Boolean searches. Stem searches detect the identified keyword and its extensions. In our case, the stem searches for 'surfboard' and 'surf-board' also found surfboards and surf-boards, surfboarder and surf-boarder, surfboarders and surf-boarders, and surfing and surf-boarding. Unfortunately, the stem searches for 'surfboard' and 'surf-board' also recognized a nationally significant horse race whose career peaked in the mid-1930s. 'Surfboard' and 'surf-board' were preferred to the other obvious keywords such as 'surf' and 'surfing' because these terms were too general and returned responses related to virtually any activity in the ocean. In order to narrow the focus of the search process to women's surfing, Boolean searches, which enable the combination of keywords, were conducted. The Boolean searches added 'female', 'girl', 'lady', 'woman' and 'women' to 'surfboard' and 'surf-board'.<sup>41</sup>

The second dimension of our distant reading of women's surfing in Australia was the visualization of the results from *Trove*. The graphical visualization was generated by an online computer application, QueryPic. At one level, QueryPic is a visualization tool that develops graphical

representations from search results in Trove; at another level, QueryPic is a product of an ongoing research venture with a fully accessible programming code.<sup>42</sup> While QueryPic illustrates the significance of creating computer-based tools for researching in digital humanities and the promise of similar applications in the future, the capacity of QueryPic to create graphs of surfboard riding and women's surfboard riding was specifically utilized. Here, search results are represented as a static visualization (Figure 2), but on QueryPic's website the graph is interactive and available for other researchers to explore the data and interrogate the conclusions.<sup>43</sup>

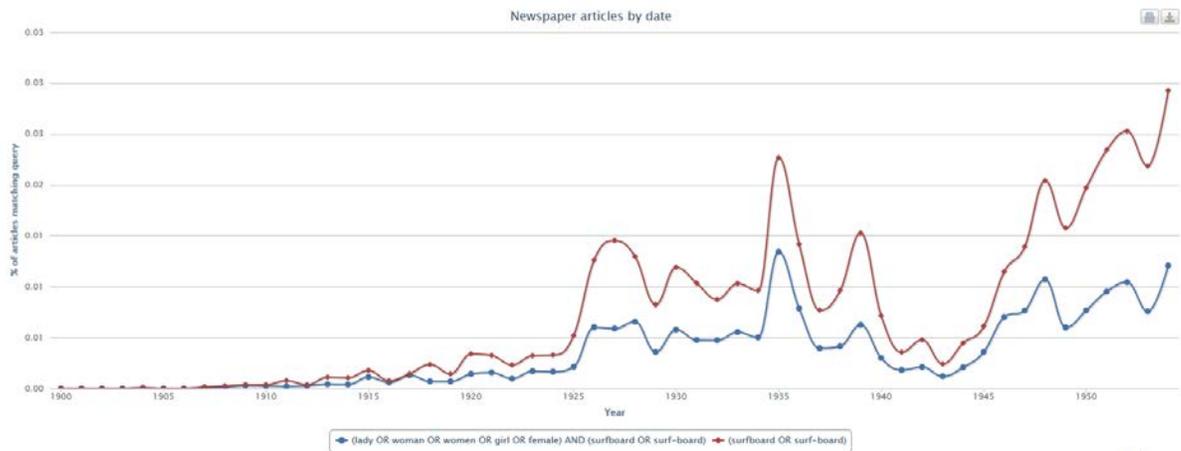


Figure 2: Percentage of articles in *Trove* that mention surfboards (1900-54)

The graph reveals the noteworthiness of surfboard riding; it does not necessarily inform us about the popularity of the activity for men or women. Noteworthiness may be related to a whole host of factors: it could be about the novelty or dangers of surfboard riding, or about international surfing, or it could be that more Australians were taking up the activity. Any one or a combination of these circumstances provides valid explanations for newspaper coverage of surfboard riding in Australia.

If the relationship between the two searches of surfboard riding and women's surfboard riding is examined, there are several obvious features. The graph of women's surfboard riding virtually mirrors

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

the representation of surfboard riding overall. There are, however, two notable exceptions. From the mid-1940s, the newsworthiness of surfboard riding increases quickly and beyond any previous level, but the coverage of women's surfboard riding does not match this trajectory. In fact, the biggest difference between surfboard riding generally and women's surfboard riding over the first six decades of the twentieth century is in 1954. The second major exception is in the second decade of the twentieth century. In this period, the noteworthiness of women's surfboard riding almost equals the coverage of surfboard riding more broadly. If our interest was in the increased newsworthiness of surfboard riding in general, distant reading would be used as a cue to pursue newspaper coverage from 1945 in a more detailed way.

Our interest, though, was in the emergence of women's surfboard riding, and distant reading points to the period of the second decade of the twentieth century when the coverage of women's surfboard riding almost equates the total coverage of surfboard riding. The second decade of the twentieth century, as indicated in the graph, is not the period of the greatest coverage of women's surfboard riding but it signifies a time-frame when the gap in coverage is the least. In many ways, this matches the stories of Letham surfing with Duke Kahanamoku, but the coverage exists before and after the 1915 demonstration on Sydney beaches and suggests the possibility of other events and individuals. At this point, the lens of distant reading was flipped to the more traditional historical method of close reading of newspapers. It is during this close reading that another woman – Isma Amor – emerged as a prominent female surfer, effectively challenging the significance of Letham as Australia's pioneering female surfer.<sup>44</sup>

*Case Study: Reading homophobia*

Homophobia is a major issue in sport, yet sport historians have been slow to research either homophobia or homosexuality.<sup>45</sup> One reason for this is the shortage of available sources that would illuminate the queer sporting past. Even today many athletes prefer to conceal their sexual identity, and in the past factors such as criminalization, social opprobrium and shame enforced that silence and ensured that few written records were left.

The absence of documentary evidence plagues historians of homosexuality, requiring researchers to search for 'snippets and shadows' among the extant records.<sup>46</sup> One obvious source of snippets is newspapers. Digitization has not only extended the scope and speed of access to historic newspapers but also has allowed searches for the occurrence and frequency of words, terms and phrases to facilitate qualitative cultural and linguistic analyses.<sup>47</sup>

But does the potential of digitization hold up for research on the intersections of sport, homosexuality and homophobia? In one way, the promise is empty: no new technological periscope will find historic examples of gay athletes if those athletes never self-identified or were unknown by others to be gay. In the vast ocean of print, closeted, historic gay athletes will sail safely undetected by our new radar. What can be detected more easily through digital tools, however, are print references that can illuminate attitudes towards homosexuality, deviant gendered identity and activity, and homophobia.

Previous research used digital searches to attempt a distant and close reading of the words 'effeminate' and 'sissy', however the emphasis was on a close reading of the digitized findings to explore how these words were applied. That exercise noted difficulties encountered with term searches and distant reading; specifically, issues around OCR and multiple word meanings.<sup>48</sup> This section will attempt a distant reading only, and of just one term, to make some preliminary comments on the possibilities, usefulness and limitations of the methodology.

The term chosen is 'poofter', which was 'one of the earliest Australian coinages in the homosexual field'.<sup>49</sup> Associated unambiguously with homophobia, 'poofter' derives from the widely-used nineteenth-century British slang word for effeminate homosexuals, 'poof' or 'puff'.<sup>50</sup> It is commonly heard in sport.<sup>51</sup> Digitized search technologies offer an opportunity to explore its uses in sporting contexts as reflected by the press. As in the study of 'effeminate' and 'sissy', we drew upon *Trove*. Unlike that study, however, which examined a subset of sporting newspapers, in this case articles in all digitized Australian newspapers were searched. While it is not the intention of this project to do a close reading, the results of such an analysis would allow for a comparison of the word's uses in various contexts and would locate sport-related uses of the term in non-sporting as well as only the sporting press.

As in the two previous case studies, this distant reading involved three initial steps: determining search terms for the topic to be researched, conducting the digital search, and attempting to visually represent the findings in a manner that would facilitate the identification and analysis of trends, patterns and possible anomalies. In the case of 'poofter', setting the terms appeared to be straightforward – search for the word itself and any common alternative spellings. Using the Boolean extension 'or', we conducted a search for <poofter OR poufter>.<sup>52</sup> This yielded 1,054 articles. This process appeared simple enough, and is exactly the approach that many researchers would take in looking for most terms. *Trove* offers a one- to two-line summary of each hit, which allows researchers to check the findings at a glance. For a researcher whose intention is to use the initial search as a guide to selecting articles that might be worth reading, these search results would be very useful. For distant reading – plotting a graph that can be considered a valuable guide to spotting usage patterns and thereby guide a closer reading – these initial search results are not necessarily trustworthy. To check, a first-order close reading, or skimming of the search results for obvious anomalies, was required.

A quick glance at the first page of results, merely 20 of the thousand-plus hits, quickly revealed anomalies. Immediately, we noticed that in one finding the word 'pound' was misidentified. Speculating that this might not be an isolated instant, we revised the search from <poofter OR poufter> to <(poofter OR poufter) NOT pound>. The original findings dropped from 1,054 to 989. A second quick scan of the screen found the word 'power' among the search results. Again, we altered the search query to eliminate this term. 828 hits remained. Alert now to gremlins, including the family name 'Poulter', we scanned for incorrect findings and constructed the following search query: <(poofter OR poufter) NOT poorer NOT potter NOT porter NOT pouffer NOT pound NOT power NOT poulter NOT Poulter>. This resulted in a near halving of the original findings, with 596 hits.

Two questions emerged: what was occurring to muddy the search, and could we reliably conduct a distant reading and graphical representation based on the 596 remaining findings? The errors were the result of issues with OCR, a common problem with digitized search findings. Many words are illegible in the original printed copy of a newspaper or in the scanned facsimile of that text. Unlike the human eye and brain, which can deduce meanings of misspelled words from their immediate placement context in a sentence or paragraph, OCR lacks such finesse.<sup>53</sup> As a result, as we glanced through the 596 findings, we located instances of the word 'poofter' where another word in the original article had been incorrectly recognized by Trove's OCR software. These included 'the Right Rev. Bishop Potter', who might roll over in his grave could he realize that he has now become 'ihe Rj-ght Kev. Bahop PoOfter'. Likewise 'Mr C. J. Booker', reappearing as 'Mr. c. J. Poofter'. Or the Coopers who have become Poofthers. One football result noted a team's 'Wieir third\* feut-poufter', which on a magnification of the digitized article itself turned out to be 'their third six-pointer'.<sup>54</sup> And in several mining reports, pit depths of, say, '300ft. Level' were recognized as 'poofter' (presumably the 300ft read as poofter). A closer analysis revealed that only two of the 72 findings before the year 1900 were actually of the word 'poofter'.

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

With such a high degree of OCR error, the findings were not suitable for a distant reading via graphic visualization. The result would have been next to useless for identifying trends and patterns: as Martin Johnes and Bob Nicholson argue on this point, 'just a few incorrect articles will skew the results'.<sup>55</sup> This does not devalue the findings themselves, however. But rather than weed out more of these misrepresentations by experimentation with additional search delimiters, the relatively small quantity of findings made it more economical to instead commence a close reading of the articles themselves. 'Bishop Poofers' and distances measured in feet could be dismissed quickly, and actual uses of the term poofter readily identified.

#### Conclusion: Distant Reading Brings us up Short

As these three case studies indicate, the benefits of distant reading are dependent on a number of factors. They all rely on the quality of the original printed matter, the digital copy of the originals and the accuracy of OCR. Faint, blurred, smudged or otherwise only partially legible text and low accuracy from the OCR, as a result of text quality or the limitations of the program itself, will always negatively affect distant reading. The attempted distant reading of homophobia in sport is a case in point. Searching for the selected term, 'poofter', was ultimately unsuccessful because the poor quality of the digital copies of newspapers resulted in a high degree of OCR error. The number of false finds was so large that under these conditions pursuing distant reading and plotting the findings on a visualization was not worthwhile. The benefits were more in the search results themselves. While the search results were cluttered with false hits, when 'weeded' they yielded many examples of the word 'poofter' across time and space. A detailed close reading of some of these located its use in a homophobic sense in 1900, prior to a previous claim that it first appeared in the Australian papers in 1903;<sup>56</sup> derogatory uses of the

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

word even earlier with meanings that may or may not have implied a homosexual slur;<sup>57</sup> cricketers reporting its use as an on-field insult in the 1920s;<sup>58</sup> and more recent uses in the context of spectator behaviours.<sup>59</sup> As valuable as these findings are, locating them was not dependent on a distant reading.

The other two case studies, however, tell a different story. These studies point to the importance of the word search function, which is the only dimension of OCR that historians have control over. The search for women's surfing was reasonably successful and produced a visualization that reflected surfing more broadly. Identifying the appropriate generic terms was possible and adding additional keywords in Boolean searches certainly helped identify some trends. These trends proved fruitful, but the process produced a considerable number of false finds, most notably the successful racehorse Surfboard, that were very difficult to systematically exclude. Nevertheless, the process of distant reading, particularly the intersection of the lines of the graph in the second decade of the twentieth century, pointed us in the direction of a new historical figure in Isma Amor.<sup>60</sup>

Distant reading was also very revealing in the case study of Muhammad Ali. A very high rate of accurate OCR recognition, created by the quality of the digital copy and the uniqueness of the names Muhammad Ali and Cassius Clay, generated a revealing graph based on the word frequency of the boxer's names. The distant reading highlighted three distinct periods that characterized the ways journalists used the two names: Cassius Clay (1964-67), both Cassius Clay and Muhammad Ali (1967-69), and Muhammad Ali (1969-). The graphical representation of the names of the boxer in the *Los Angeles Sentinel* displayed trends and patterns that prompt hypotheses and questions about the relationship between the media's coverage of the boxer and racial, religious and political issues during the 1960s and 1970s.

What we take from these examples, given the current level of development in historical computer-reading programs, is that distant reading has some incredibly valuable virtues but is rarely an

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

end in itself. In the future, when these computing reading programs become more sophisticated, distant reading will take historians in new directions and could be an end in itself. However, as Sculley and Pasanek argue and as evidenced by these case studies, contemporary automated analysis still brings 'us up short'.<sup>61</sup> While distant reading helps shape the historical task by enabling big picture analysis, encouraging different questions, and forming new hypotheses, it does not complete the totality of the historical process. Distant reading of newspapers enables us to create a road map that provides multiple ways not only to get to destinations but also to find new destinations. But if we want to fully explore both the journey *and* the destination, then this will require the close reading skills that historians have always embraced in representing the past.

Both the work on Ali and women's surfing highlight the value of combining distant and close reading in the history-making process. In the case study of Ali, distant reading suggests lines of enquiry, yet it is only through close reading that it is possible to comprehend the complexities and nuances that are evident in the responses of the black journalists who wrote for the *Sentinel*. Close reading is necessary to place the *Sentinel* within the broader context of the Black Press in the 1960s and 1970s; to investigate the unique culture of Los Angeles and make suggestions as to how journalists from that city might interpret Ali in ways that are different from journalists from Chicago, or Baltimore, or New York. It is only through close reading that we can we start to address important questions such as: why was the *Sentinel*, as an outlet for, and creator of, African American opinion, reluctant to acknowledge the boxer's name change in 1964?; what changes occurred post-1967 that rationalized the use of both Cassius Clay and Muhammed Ali by the *Sentinel* journalists?; and why did the name Muhammad Ali become widely accepted from 1969/1970 and what did that mean for the African American community in Los Angeles? Answering these questions through close reading is the next step in the process of understanding how shifting usage of the boxer's names related to discourses about politics, race and religion.

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

Similarly, the value of using distant and close reading is demonstrated in the example of Australian women's surfing. Distant reading highlighted a particular period of surfing, but it was the close reading of digitized newspaper articles of this period that was most revealing. The close reading task was approached with a strong expectation that one woman would feature prominently: Isabel Letham. Letham dominates the popular memory of Australian women's surfing because she rode tandem with Duke Paoa Kahanamoku, the visiting Hawaiian Olympic swimming champion, for several intervals during an hour-long exhibition at Dee Why Beach in 1915. This ride is folkloric in Australian surfing, celebrated across a range of social memory sites and recently the subject of a book romanticizing the relationship between Letham and Kahanamoku.<sup>62</sup> Letham is frequently credited as Australia's first surfboard rider and only female Australian surfer of the era. A close reading of the digitized newspapers, and the kind of 'marinating' that Darnton referred to in the non-digitized archives in Sydney, identified a new identity, Isma Amor. Close reading facilitated a content, semiotic and material analysis of a seminal surfing photograph, helped to analyze the reasons for the marginalization of Amor's contribution to surfing, and ultimately contributed to deconstructing the Letham/Kahanamoku myth by placing Amor at the centre of emergence of women's surfing in Australia.<sup>63</sup>

The three case studies in this article reinforce a larger debate and commentary about distant reading in the digital humanities. In essence, this argument recognizes how the 'infinite' archive, the development of computer reading machines and new forms of evaluating big data have altered the cognitive horizons of researchers in the digital era. Furthermore, there are very few in the humanities who shun close reading in favour of an exclusive focus on distant reading. Instead there is a very clear agenda to utilize distant and close reading in tandem or, as Jones envisions, the creation of a hermeneutic circle between the two forms of analysis.<sup>64</sup> As Burdick and colleagues argue: 'The digital

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

humanist is capable of "toggling" between views of the data, zooming in and out, searching for large-scale patterns and then focussing in on fine-grained analysis'.<sup>65</sup> This 'toggling' is a very accurate description of the research process as we pivoted between distant and close readings to examine Muhammad Ali, women's surfing and homophobia.

## Notes

---

<sup>1</sup> Gary Osmond and Murray G Phillips (eds), *Sport History in the Digital Era* (Urbana: University of Illinois Press, 2015).

<sup>2</sup> John Theibault, 'Visualizations and Historical Arguments', in Kristen Nawrotzki and Jack Dougherty (eds), *Writing History in the Digital Age* (Ann Arbor: University of Michigan Press, 2013), 173-85.

<sup>3</sup> James Goulding, 'Historical Thinking and Social Media', *Agora* 46, no. 3 (2011), 11-19.

<sup>4</sup> Lev Manovich, 'Trending: The Promises and the Challenges of Big Social Data', in Matthew K. Gold (ed.), *Debates in the Digital Humanities* (Minneapolis: University of Minnesota Press, 2012), 460-75.

<sup>5</sup> For an extended discussion of the tripartite model of history making in the digital era, see Osmond and Phillips, *Sport History in the Digital Era*.

<sup>6</sup> Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner and Jeffrey Schnapp, *Digital Humanities* (Cambridge, MA: MIT Press, 2012), 37.

<sup>7</sup> Paul Turnbull, 'Historians, Computing and the World-Wide-Web', *Australian Historical Studies* 41, no. 2 (2010): 131-48.

<sup>8</sup> Wray Vamplew, 'The Power of Numbers: A Plea for More Quantitative Sports History', *Sport in Society: Cultures, Commerce, Media, Politics* (2015): DOI: 10.1080/17430437.2015.1056574

<sup>9</sup> Alun Munslow, *Narrative and History* (Houndmills: Palgrave Macmillan, 2006).

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

---

<sup>10</sup> Alun Munslow, *The Future of History* (Houndmills: Palgrave Macmillan, 2010).

<sup>11</sup> For discussion of these issues by sport historians see Douglas Booth, *The Field: Truth and Fiction in Sport History* (Abington: Routledge, 2005); Murray G. Phillips, 'Deconstructing Sport History: The Postmodern Challenge', *Journal of Sport History* 28, no. 3 (2001), 327-45; Murray G. Phillips, (ed.), *Deconstructing Sport History: A Postmodern Analysis* (New York: SUNY, 2006) and Richard Pringle and Murray G Phillips (eds), *Examining Sport Histories: Power, Paradigms, and Reflexivity* (Morgantown, WV: FIT, 2013).

<sup>12</sup> Burdick, et al., *Digital Humanities*, 37.

<sup>13</sup> David M. Berry, 'Introduction: Understanding Digital Humanities', in David M Berry (ed.), *Understanding Digital Humanities* (Houndmills: Palgrave Macmillan, 2012), 2.

<sup>14</sup> David Thomas and Valerie Johnson, "'New Universes or Black Holes?' Does Digital Change Anything", in Toni Weller (ed.), *History in the Digital Age* (London: Routledge, 2013), 173-93.

<sup>15</sup> Foucault's and Derrida's analyses of the archive are discussed and expanded in Carolyn Steedman, *Dust: The Archive and Cultural History* (New Brunswick, NJ: Rutgers University Press, 2001).

<sup>16</sup> Carolyn Hamilton, Verne Harris, Jane Taylor, Michele Pickover, Graeme Reid and Razia Saleh (eds), *Refiguring the Archive* (Dordrecht: Kluwer Academic, 2002).

<sup>17</sup> Robert Darnton, 'The Good Way to do History', *New York Review of Books*, 9 January 2014, 52.

<sup>18</sup> Arlette Farge, *The Allure of the Archives*, trans. Thomas Scott-Railton (New Haven, CT: Yale University Press, 2013).

<sup>19</sup> Leighton Evans and Sian Rees, 'An Interpretation of Digital Humanities', in David M Berry (ed.), *Understanding Digital Humanities* (Houndmills: Palgrave Macmillan, 2012), 22.

<sup>20</sup> Bob Nicholson, 'Counting Culture; or, How to Read Victorian Newspapers from a Distance', *Journal of Victorian Culture* 17, no. 2 (2012), 246.

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

---

<sup>21</sup> Bob Nicholson, 'The Digital Turn: Exploring the Methodological Possibilities of Digital Newspaper Archives', *Media History* 19, no. 1 (2013), 67.

<sup>22</sup> Matthew L. Jockers, *Macroanalysis: Digital Methods and Literary Theory* (Urbana: University of Illinois Press, 2013), 171 makes this point about literary studies but we think it applies equally to historical research.

<sup>23</sup> Caroline Daley, 'Papers from the Past and Problems from the Present', *Turnbull Library Record* 43 (2010), 65-72.

<sup>24</sup> Paul Hagon, 'Trove Crowdsourcing Behaviour', National Library of Australia, [http://www.information-online.com.au/pdf/Tuesday\\_Concurrent\\_2\\_1125\\_Hagon.pdf](http://www.information-online.com.au/pdf/Tuesday_Concurrent_2_1125_Hagon.pdf) (accessed 19 October 2013).

<sup>25</sup> Jim Mussell, 'Doing and Making: History as Digital Practice', in Toni Weller (ed.), *History in the Digital Age* (London: Routledge, 2013), 61-3.

<sup>26</sup> Franco Moretti, *Graphs, Maps, Trees: Abstract Models for a Literary History* (London: Verso, 2005).

<sup>27</sup> Franco Moretti, *Distant Reading* (New York: Verso, 2013).

<sup>28</sup> Nicholson, 'Counting Culture', 243-44; Dallas Liddle, 'Reflections on 20,000 Victorian Newspapers: "Distant Reading" *The Times* Using *The Times Digital Archive*', *Journal of Victorian Culture* 17, no. 2 (2012): 230-37.

<sup>29</sup> For a more extended discussion of this issue see Burdick, et al., *Digital Humanities*, 103-106; Alan Liu, 'Where Is Cultural Criticism in the Digital Humanities?' in Gold (ed.), *Debates in the Digital Humanities*, 490-509; Berry, 'Introduction'; Katherine N. Hayles, 'How We Think: Transforming Power and Digital Technologies', in Berry (ed.), *Understanding Digital Humanities*, 42-66.

<sup>30</sup> The Black Press underwent a number of evolutions throughout the nineteenth and twentieth centuries. However, most publications, including the *Los Angeles Sentinel*, maintained their focus on

black issues and black advocacy. Understandably, during the 1960s this focus was grounded in the Civil Rights Struggle. For a more complex exploration of the Black Press during the Civil Rights Era, see Lawrence A. Eldridge, *Chronicles of a Two-Front War: Civil Rights and Vietnam in the African American Press* (Columbia: University of Missouri Press, 2011).

<sup>31</sup> We would like to point out the serendipity of this situation. The purchase of a portion of ProQuest's *Historical Newspapers* was initially refused by our university library because of the high costs, but was eventually made because there were surplus funds at the end of the financial year that needed to be spent. The project on Muhammad Ali could not have been conducted without this purchase.

<sup>32</sup> See ProQuest Historical Newspapers Pamphlet, <http://media2.proquest.com/documents/Historical+Newspaper+Brochure.pdf>, (accessed 18 August, 2015).

<sup>33</sup> Despite their distaste for the name, members of the Nation of Islam were often referred to as 'Black Muslims' by the press, by virtue of the group's almost exclusively African-American membership. See David Remnick, *King of the World: Muhammad Ali and the Rise of an American Hero* (London: Picador, 2000), 207-8.

<sup>34</sup> Mike Marqusee, 'Sport and Stereotype: From Role Model to Muhammad Ali', *Race and Class* 36, no.4 (1995), 15.

<sup>35</sup> Geographically, California is not a Northern state. However, in the context of the Civil Rights Movement, California might be seen as 'Northern' in its sensibilities toward black liberation and integration.

<sup>36</sup> See Eldridge, *Chronicles of a Two-Front War* for a more comprehensive discussion of the Black Press during the Vietnam War/Civil Rights era. See also: Daniel Bennett Coy, 'Imagining Dissent: Muhammad Ali, Daily Newspapers, and the State.' Master's Thesis, University of Tennessee Knoxville (2004), 6; Jack

Nelson, 'The Civil Rights Movement: A Press Perspective.' *Human Rights* 28, no. 4 (2001), 3-6 for a discussion of 'white' newspaper reactions to the Civil Rights Movement, the Nation of Islam, and the black liberation movement more broadly.

<sup>37</sup> Douglas Booth names several women who were active participants in the nascent Sydney surfing culture of the early nineteenth hundreds; Leonie Huntsman describes the formation of and resistance to women's lifesaving clubs from as early as 1908; and Ed Jaggard contends that women competed nationally in surf lifesaving from the 1920s onwards: see Douglas Booth, *Australian Beach Cultures: The History of Sun, Sand and Surf* (London: Frank Cass, 2001), 102; Leone Huntsman, *Sand in Our Souls: The Beach in Australian History* (Melbourne: Melbourne University Press, 2001), 94–96; and Ed Jaggard, "'Tempering the Testosterone": Masculinity, Women and Australian Surf Lifesaving', *The International Journal of the History of Sport* 18, no. 4 (2001), 17.

<sup>38</sup> Gary Osmond, 'Myth-Making in Australian Sport History: Re-Evaluating Duke Kahanamoku's Contribution to Surfing', *Australian Historical Studies* 42, no. 2 (June 2011), 260–76.

<sup>39</sup> For a rare profile of a female board rider, Grace Smith Wootton, who began surfing in Victoria in 1915, see Lana Wells, *Sunny Memories: Australians at the Seaside* (Melbourne: Greenhouse, 1982), 156–59.

<sup>40</sup> The exceptions to copyright limitations are the *Canberra Times* and *Australian Women's Weekly*.

<sup>41</sup> <http://dhistory.org/querypic/d6/> (accessed 6 November 2014).

<sup>42</sup> 'QueryPic', <http://dhistory.org/querypic/> (accessed 6 November 2014).

<sup>43</sup> 'Surfing and Female Surfing in Australia (Graph 1)', <http://dhistory.org/querypic/d6/> (accessed 6 November 2014).

<sup>44</sup> For an extended discussion of this project, see Murray G. Phillips and Gary Osmond, 'Australia's Women Surfers: History, Methodology and the Digital Humanities', *Australian Historical Studies* 46, no. 2 (2015), 285-303.

<sup>45</sup> See Erik Denison and Alistair Kitchen, 'Out on the Fields: The First International Study on Homophobia in Sports', 2015, <http://www.outonthefields.com/> (accessed 7 August 2015).

<sup>46</sup> Lisa Featherstone, 'Snippets and Shadows of Stories: Thoughts on Sources and Methods when Writing an Australian History of Sexuality', in Graham Willett and Yorick Smaal (eds), *Intimacy, Violence and Activism: Gay and Lesbian Perspectives on Australasian History and Society* (Melbourne: Monash University Publishing, 2013), 80.

<sup>47</sup> For a pioneering study, see: Jean-Baptiste Michel, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, et al., 'Quantitative Analysis of Culture Using Millions of Digitized Books', *Science* 331, no. 6014 (January 14 2011): 176-82.

<sup>48</sup> Gary Osmond, "'Pink Tea and Sissy Boys": Digitized Fragments of Male Homosexuality, Non-heteronormativity and Homophobia in the Australian Sporting Press, 1845-1954', *The International Journal of the History of Sport* (2015 in press).

<sup>49</sup> Gary Simes, 'The Language of Homosexuality in Australia', in Robert Aldrich and Garry Wotherspoon (eds), *Gay Perspectives: Essays in Australian Gay Culture* (Sydney: University of Sydney, 1992), 41.

<sup>50</sup> Ibid.

<sup>51</sup> Note, for example, the reflexive title of Johnny Warren's book, *Sheilas, Wogs & Poofers: An Incomplete Biography of Johnny Warren and Soccer in Australia* (Sydney: Random House Australia, 2002).

<sup>52</sup> This could be extended to include other spelling variants, like 'poof' and 'poofa', but the aim of this case study was to test findings rather than attempt to be exhaustive.

<sup>53</sup> For a discussion of the limitations of OCR recognition, see Martin Johnes and Bob Nicholson, 'Sport History and Digital Archives in Practice', in Gary Osmond and Murray G. Phillips, *Sport History in the Digital Era* (Urbana, University of Illinois Press, 2015), 56-57. For factors that influence OCR accuracy in

Murray G. Phillips, Gary Osmond and Stephen Townsend, "A Bird's-eye View of the Past: Digital History, Distant Reading and Sport History". *The International Journal of the History of Sport*. Published online 28 October 2015. DOI: 10.1080/09523367.2015.1090976

---

historic newspapers, see Rose Holley, 'How Good Can It Get?: Analysing and Improving OCR Accuracy in Large Scale Historic Newspaper Digitisation Programs.' *D-Lib Magazine*. March/April 2009, vol. 15 no 3/4. <http://www.dlib.org/dlib/march09/holley/03holley.html> (accessed 4 August 2015).

<sup>54</sup> *Mirror* (Perth), 11 September 1937, 4,

[http://trove.nla.gov.au/ndp/del/article/75451933?searchTerm=%22Wieir%20third\\*%20feut-poufter%22&searchLimits=exactPhrase=Wieir+third\\*+feut-](http://trove.nla.gov.au/ndp/del/article/75451933?searchTerm=%22Wieir%20third*%20feut-poufter%22&searchLimits=exactPhrase=Wieir+third*+feut-poufter)

[poufter](http://trove.nla.gov.au/ndp/del/article/75451933?searchTerm=%22Wieir%20third*%20feut-poufter) |||anyWords|||notWords|||requestHandler|||dateFrom|||dateTo|||sortBy. Correcting the text does not change the search results: although we amended this line in the article, which now appears correctly in the 'Electronically Translated Text' panel on *Trove* that accompanies the digitized scan of the article, it continues to appear in a search for 'poofter'.

<sup>55</sup> Johnes and Nicholson, 'Sport History and Digital Archives in Practice', 63.

<sup>56</sup> *Truth*, 25 March 1900, 5; Simes, 'The Language of Homosexuality in Australia', 41.

<sup>57</sup> *North Melbourne Advertiser*, 19 January 1889, 2.

<sup>58</sup> *Geelong Advertiser*, 21 December 1921, 3.

<sup>59</sup> *Canberra Times*, 25 June 1980, 21.

<sup>60</sup> Phillips and Osmond, 'Australia's Women Surfers'.

<sup>61</sup> D. Sculley and Bradley M. Pasanek, 'Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities', *Literary and Linguistic Computing* 23, no. 4 (2008), 423.

<sup>62</sup> Phil Jarratt, *That Summer at Boomerang* (Melbourne: Hardie Grant, 2014).

<sup>63</sup> Phillips and Osmond, 'Australia's Women Surfers'.

<sup>64</sup> Steven E. Jones, *The Emergence of Digital Humanities* (New York: Routledge, 2014).

<sup>65</sup> Burdick, et al., *Digital Humanities*, 39.