3

THE IRISH WORLD: HOW TO REVISE A LONG-STANDING DICTIONARY PROJECT

TURLOUGH O'RIORDAN

The gradual emergence of digital resources since the 1980s has transformed biographical research. The proliferation of digital resources—ubiquitous and searchable—generates profound questions for the integrity of biographical research. The digitisation of existing scholarly material (monographs, edited collections, journal articles) and new digital platforms (newspaper archives, online repositories), when combined with digital search techniques, facilitate the querying of vast amounts of information. The purpose of biographical reference remains unaltered. Offering accuracy and precision, national biographical dictionaries present rigorously researched factual biographies, incorporating relevant contextual analysis, based on verifiable sources.

Although perceived in analogue terms, national biographical dictionaries are expected to incorporate the outputs of the digital revolution. The increasing prevalence of digitised primary resources—newspapers and periodicals; official and public papers; specialised archives and collections— has transformed historical research practices. Digital platforms, editions, and tools deliver an ever-expanding abundance of information. However, established qualitative research methods, developed in analogue contexts, are unquestionably applied to digital material and resources.

Digital publication allows new content to be rapidly disseminated and easily searched. Scholarship assesses the veracity and authenticity of information. It must also consider how such information is (or is not) presented to us. Fluid conceptual boundaries between digital editions, archives, and online platforms[1] (all structured databases) are blurred by how keyword searching guides scholarship. Our unquestioned reliance upon these tools, whose operation is little understood, is troubling. The experience of the *Dictionary of Irish Biography* (*DIB*), facing into a world increasingly reliant upon digitally mediated research, may be instructive.

The *Dictionary of Irish Biography*

The *DIB* commenced in spring 1983. Preliminary scoping work by the Royal Irish Academy focused on trawling sources and consulting widely amongst experts. As initially conceived, external specialists would research lives from defined periods, to be collected and periodically published. Editorial planning became orientated towards developing a database to be published as a CD-ROM. James McGuire was appointed as managing editor in the early 1990s. A publishing contract with Cambridge University Press, agreed in 1997, was soon followed by significant funding from the Irish Higher Education Authority. An enlarged in-house research staff of specialist scholars had, by 2003, completed around 70 per cent of the planned *Dictionary*; it was decided that year to publish simultaneously in both hard copy and online. Completed in the spring of 2009, the *Dictionary* was published that November by Cambridge University Press and the Royal Irish Academy. Launched in Dublin by An Taoiseach (prime minister) Brian Cowen, the hard copy—over 8 million words spread across 10,000 pages and nine volumes—then mirrored exactly the online platform, having been conceived and executed as a hard-copy reference project.

---

1   Kenneth M. Price, 'Edition, Project, Database, Archive, Thematic Research Collection: What's in a Name?', *DHQ: Digital Humanities Quarterly* 3, no. 3 (2009), accessed 2 November 2016, digitalhumanities.org:8081/dhq/vol/3/3/000053/000053.html. *RIDE: A Review Journal for Digital Editions and Resources* (ride.i-d-e.de) seeks to redress the extent to which digital editions and platforms are ignored by established journals, subjecting them to rigorous scholarly peer-review.

Emphasising factual accuracy, embracing modern scholarship, and seeking to ensure accessibility to the general reader, the *DIB* is a general biographical reference work.[2] The 2009 edition comprises individuals who died before the end of 2002, either born in Ireland (with careers within or outside Ireland) or elsewhere (with careers in Ireland). The 'essential criterion of selection is presence in Ireland'.[3] The preponderance of military, ecclesiastical, judicial, and political subjects in earlier Irish biographical dictionaries saw the editors give 'primacy to achievement over position', seeking to include 'those names which seem most likely to be the objects of enquiry in the twenty-first century'.[4]

Encompassing mythical figures such as the tragic Deirdre 'of the sorrows', the legendary warrior Fionn mac Cumhaill, and Molly Malone (the figurative Dublin street hawker), all likely euhemerised, the *DIB* includes saints—such as Finbarr of Cork and of course St Patrick—and the medieval lodestar Gerald of Wales, author of the *Typographia Hibernia*. Arkle, the famous horse, could not be included, though his owner Anne, Duchess of Westminster, is given extensive treatment. Questions of Irishness are resolved by treating the island of Ireland from the earliest times to the present day.

The 9,700 lives (in 9,014 signed entries) comprising the 2009 first edition were presented identically to readers in nine volumes and online. From late 2009, the *DIB* began adding supplementary lives (biannually in June and December); individuals are considered for inclusion six years after their death. Gradually, we commenced revising existing entries and addressing corrigenda (biannually each spring and autumn). Errors of fact and interpretation are rectified, guided by our own research and external submissions. Assessing new sources and evaluating existing findings is an important, though resource-intensive, part of our work. There has been a gradual bifurcation between the static, analogue, hard-copy dictionary and the iterative online digital platform.[5]

---

2   What constitutes a 'general reader' is of course open to debate, and has been the source of some of the most enjoyable discussions with colleagues on the *DIB* over the last 18 years.

3   James McGuire and James Quinn, 'Introduction', in *Dictionary of Irish Biography*, eds James McGuire and James Quinn, vol. 1, *A–Burchill* (Cambridge: Cambridge University Press, 2009), xi.

4   McGuire and Quinn, 'Introduction', xxi.

5   The June 2018 supplemental update to the *DIB* online brings the total number of lives to 10,461. In August 2018, hard-copy volumes 10 and 11 of the *DIB*, comprising those who died between 2003 and 2010 alongside the first two batches of 'missing persons', were published by Cambridge University Press.

Editorial Considerations

Assessment of the recently dead requires sensitivity. The absence of accessible primary source material led to many such entries being written by contemporaries who knew those treated, or had notable insight into their life and times. Issues of objectivity and sensitivity also impinge upon supplementary lives added to the *DIB* since 2009, especially those dying in contentious or tragic circumstances, or who were involved in traumatic events.

Consultation amongst experts yields a viable collection of subjects for consideration. Blind spots can be rectified by the later identification and the inclusion of 'missing persons'; 82 such lives were added to the *DIB* in December 2013, 75 in December 2015, and a further 57 in December 2017. Historiographical developments guide selection, aided by newly digitised sources being fully text-searchable.

Questions of representation and coverage (reliant on subjective assessments) reflect the wider composition of society and its expectations. Newspaper obituaries, though sometimes unreliable, guide initial identification. However, those who have emigrated, or died long after their peer group, may suffer from exclusion bias in the reporting of Irish newspapers and the publications of professional and voluntary associations. Certain careers—for example, engineering, medicine, and science—are increasingly professionally introverted. The specialisation upon which professional achievement is based means notable individuals are often little known beyond their specialist fields.

The Irish World

Periods of sustained emigration from Ireland spur transnational lives, which are easily overlooked. David Orr (1922–2008), a leading United Kingdom businessman, was of some prominence in Ireland, while Arthur Gwynn (1908–2008), a medic, naturalist, and explorer in Australia, was little known in Ireland and less so in Australia.[6] Orr's public profile

---

6 Turlough O'Riordan, 'Arthur Montagu Gwynn (1908–2008)', and 'David Alexander Orr (1922–2008)', *Dictionary of Irish Biography*, June 2014, accessed 9 August 2017, dib.cambridge.org.

(a British soldier awarded the prestigious Military Cross, who became a leading captain of industry), and to a lesser extent Gwynn's, being a scion of a prominent Irish scholarly family, brought each to our attention.

Bryan Beirne (1918–1998), a world-renowned entomologist in Canada, emigrated there after a starred university career in Dublin. Jeremy Swan (1922–2005), a prominent cardiologist in the United States of America, and John Crofton (1912–2009), a tuberculosis physician and public health expert in the United Kingdom, followed similar paths.[7] Though little known in Ireland, each reached the apex of their specialist fields overseas, their emigration spurred by a dearth of postwar opportunities in Ireland. Crofton and international relations scholar Fred Halliday (1946–2009) both compiled *Who's Who* entries. Beirne and political scientist Peter Mair (1951–2011) both left published *curricula vitae*, valuable for compiling professional chronologies for those active in academia and the sciences. Halliday, born in Ireland, spent most of his life in the United Kingdom. Mair, educated in Ireland, emigrated as a postgraduate and rose to prominence in Europe.[8] Irish people with overseas careers are often harder to identify and thus less likely to be considered for inclusion. Unfortunately emigration from Ireland continues, with 480,000 people emigrating from Ireland between 2008 and 2014, just under 10 per cent of our population.[9]

It is significantly easier to identify figures coming to Ireland. Heinrich Böll (1917–1985), the German writer, spent time in Ireland during the 1950s and 1960s. His *Irisches Tagebuch* (*Irish Journal*), published in 1957 and translated into English in 1967, was influential in both countries. Ivan Beshoff, 'Russian immigrant and purveyor of fish and chips',[10] was a mutineer of the Russian battleship *Potemkin,* introduced to Lenin in London by the Irish labour leader James Larkin (1874–1947).[11] They

---

7   Turlough O'Riordan, 'Bryan Patrick Beirne (1908–1998)', 'Jeremy Swan (1922–2005)', and 'John Wenman Crofton (1912–2009)', *Dictionary of Irish Biography*, 2011–16, accessed 9 August 2017, dib.cambridge.org.

8   Turlough O'Riordan, 'Fred Haillday (1946–2009)', and 'Peter Mair (1951–2011)', *Dictionary of Irish Biography*, June and December 2016, accessed 9 August 2017, dib.cambridge.org.

9   Tom Healy, 'Emigration Has Taken Its Toll', Nevin Economic Research Institute, 3 July 2015, accessed 3 July 2017, www.nerinstitute.net/blog/2015/07/03/emigration-has-taken-its-toll/. This draws on (Irish) Central Statistics Office data, accessed 3 July 2017, www.cso.ie/en/statistics/population/archive/.

10   C. J. Woods, 'Ivan Beshoff (1882?–1987)', *Dictionary of Irish Biography*, 2009, accessed August 2017, dib.cambridge.org.

11   Emmet O'Connor, 'James Larkin (1874–1947)', *Dictionary of Irish Biography*, 2009, accessed August 2017, dib.cambridge.org.

may be contrasted with Erwin Shrödinger (1887–1961) and George Friedrich Handel (1685–1759), included due to their brief stays in Ireland.[12]

Many 'missing persons' added to the *DIB* since 2009 had previously overlooked transnational lives. Sara Blomfield (1859–1939), a Bahá'í pioneer and humanitarian in the United Kingdom; Joanna Hiffernan (c. 1843–post 1903?), an artist's model and muse of James Whistler in Paris; Tony Mullane (1859–1944), a baseball player in nineteenth-century America; William Desmond Taylor (1877–1922), a silent film director in America and celebrity murder victim; James Dalton (1834–1919), a pastoralist and merchant in Australia; and Mary Latchford Kingsmill Jones (1877–1968), a public representative in the UK, are amongst a selection added to the *DIB* since 2009.[13] Those emigrating to non-Anglophone countries are less likely to come to our attention. Jennifer Musa (1917–2008), politician and Pushtun tribal elder in Balochistan, Pakistan, and James Skinner (1923–2008), politician and lawyer in Zambia and Malawi, are amongst many fascinating supplementary transnational lives to be found in the *DIB*.[14] Musa, working as a nurse in Oxford, England, met and married the scion of a prominent Pushtun family, returning to Balochistan with him. Skinner, after training as a lawyer and facing bleak employment prospects in postwar Ireland, emigrated to the Federation of Rhodesia and Nyasaland.

---

12 Paul Collins, 'George Friedrich Handel (1685–1759)' and Patricia M. Byrne, 'Erwin Shrödinger (1881–1961)', *Dictionary of Irish Biography*, 2009, accessed August 2017, dib.cambridge.org.
13 Turlough O'Riordan, 'Sara Louise Blomfield (1859–1939)', Lawrence William White, 'Joanne Hiffernan (*c*.1843–*p*.1903?)' and 'Anthony John Mullane (1859–1944)', Patrick Maume, 'William Desmond Taylor 1872–1922)', Derek Barry, 'James Dalton (1834–1919)' and Arnold Horner, 'Mary Latchford (Kingsmill) Jones (1877–1968)', *Dictionary of Irish Biography*, 2013–15, accessed August 2017, dib.cambridge.org.
14 Turlough O'Riordan, 'Jennifer Musa (1917–2008)' and 'James John Skinner (1923–2008)', *Dictionary of Irish Biography*, June and December 2014, accessed August 2017, dib.cambridge.org.

Digital Research

Knowledge of Musa and Skinner's lives came from Anglophone sources. The cultural and linguistic affinities shared amongst postcolonial states conditioned their emigration. Are those who emigrate to non-Anglophone cultures as likely to come to our attention? Researching less prominent individuals in culturally, scientifically, or politically important though unheralded sectors (for instance, the nontraditional arts, central banking, medicine, and scientific research) benefits from digital research methods. The digitisation of material covering their lives—easily searchable and accessible—aids initial identification and subsequent research. Transnational and marginal lives are now more likely to receive due consideration and recognition. That digital resources can be easily searched undoubtedly furthers biographical research into less prominent lives.

Entry length can also be problematic, implicitly suggestive of relative importance.[15] While any subjective judgment will vary over time, national biographical dictionaries should arguably strive most to account for the lives of lesser-known individuals. Their contributions to specialist fields or less prominent sections of society,[16] likely given little coverage elsewhere, are arguably most deserving of detailed treatment. Those not treated in standard narratives due to their marginality, or given only passing mention, deserve our greatest efforts. The digital revolution greatly serves this goal: aiding the inclusion of 'missing persons' and facilitating the researching of their exploits in ways that would have been until recently unimaginable. Yet the ability to search ever more widely and deeply also generates demands for the revision or expansion of existing entries.

Beyond rectifying egregious errors and omissions, there is no obvious rubric guiding the incorporation of new information, sources, or research findings. Digital publication is provisional due to the ability to amend what is presented to readers, which subtly alters their expectations. There may even be a case for considering excluding individuals from the *DIB* as their historical worth is re-evaluated. If the practice of revision incorporates 'who to include', should it not concomitantly assess 'who to exclude'?

---

15   *DIB* entries range from circa 200 words to over 9,000 words for major figures, such as Éamon De Valera (1882–1975), St Patrick (c. 420–490?), and W. B. Yeats (1865–1935).

16   Discussion with colleagues has yet to yield a positive, non-pejorative term for such figures.

Revision

Lives from the eighteenth century onwards are most likely to undergo revision as relevant sources are digitised. A coterie of late twentieth-century figures, subject to various forms of public inquiry during and after their deaths, have necessitated revision.[17] James Gogarty (1917–2005), an engineer and building company executive, gave testimony into political and planning corruption that laid bare squalid malfeasance.

> [Gogarty] opened a Pandora's box of alleged corruption and cronyism, regarding the probity of the planning process, and monies received under questionable circumstances by elected politicians at local and national level, and by appointed officials. The planning tribunal launched on the basis of Gogarty's allegations sat for fifteen years (1997–2012), the longest tribunal of inquiry in the history of the Irish state.[18]

Subsequent complex legal proceedings resulted in the January 2015 withdrawal of many findings of fact against named individuals based on Gogarty's evidence. This in turn required the revision of *DIB* entries treating associated business and political figures. Two medics*,* Margaret Dunlevy (1909–2002) and Patrick Meenan (1917–2008), were variably involved in vaccine trials undertaken in Irish children's homes in the 1960s and 1970s.[19] Legal proceedings concluded in 2003, and 2015 public revelations, impacted how the historical record construes their careers. Emergent legal, or more likely historical, findings will require the future revision of relevant *DIB* entries. The key historiographical issue is what constitutes a reasonable assessment of the known facts at a given point in time.

On Shelves and Screens

Factual accuracy and robust interpretation, the core functions of biographical reference, require pertinent new sources to be reviewed. A balance must be struck between accuracy on the one hand, and

---

17   'Tribunals of enquiry' and other statutory forms of public investigation are common in addressing political, religious, and judicial scandals in Irish public life.

18   Lawrence William White, 'James Martin Gogarty (1917–2005)', *Dictionary of Irish Biography*, revised April 2015, accessed November 2016, dib.cambridge.org.

19   Turlough O'Riordan, 'Patrick Meenan (1917–2008)' and 'Margaret Dunlevy (1909–2002)', *Dictionary of Irish Biography*, June and December 2015, accessed November 2016, dib.cambridge.org.

immediacy and completeness on the other. Colin Allen, Uri Nodelman, and Edward Zalta posit how online reference works face an impossible triad: between 'authoritative', 'comprehensive', and 'up to date'; only two are possible.[20] Humanities scholarship demands the rigorous discussion and attribution of evidence. These scholarly editorial practices are what fundamentally distinguish academic reference works from Wikipedia and other crowd-sourced platforms. Yet Wikipedia, John Naughton argues, has developed 'a remarkable system of collective governance that combines transparency, a respect for neutrality and pragmatism'.[21]

Humanities scholarship interprets evidence in pursuit of rigorous discussion in contrast to the binary analyses found in the pure and applied sciences. The objective balancing of evidence against scientific theory may better suit the communal, mediated nature of Wikipedia. Inadequate when a topic demands critical appraisal (humanities and social science), useful when a topic is centred on uncontested factual information (pure and applied sciences), Wikipedia's 'increasing ubiquity as a source of information' risks engendering widespread reliance upon a single source, what Jack Lynch terms an 'information monoculture'.[22] The reinforcing nature of online search promotes open-access content: the more people search for and use a resource, the more it is returned in search results, the more it is used, and so on. Yet Wikipedia's popularity, magnified by its noble open-access model, should not be maligned. The quality and utility of many of its articles, biographical or otherwise, was praised by Roy Rosenzweig in a seminal 2006 article arguing the value of 'open-source' history could no longer be ignored by the academy.[23]

The experience of other academic reference works, with various organisational and funding models, may be instructive. The *Oxford English Dictionary* decided in 2012 to discontinue hard-copy publication

20 Colin Allen, Uri Nodelman, and Edward N. Zalta, 'The Stanford Encyclopedia of Philosophy: A Developed Dynamic Reference Work', *Metaphilosophy* 33, nos 1–2 (2002): 210–28. doi.org/10.1111/1467-9973.00225.

21 John Naughton, 'Jimmy Wales Goes After the Truth. Brave Man', *Guardian*, 30 April 2017, accessed 9 August 2017, www.theguardian.com/commentisfree/2017/apr/30/jimmy-wales-goes-after-truth-brave-man-wikipedia.

22 Peter Thoneman, 'The All-Conquering Wikipedia?', *Times Literary Supplement*, 25 May 2016, accessed 29 November 2016, www.the-tls.co.uk/articles/public/encyclopedic-knowledge.

23 Roy Rosenzweig, 'Can History Be Open Source? Wikipedia and the Future of the Past', *Journal of American History* 93, no. 1 (2006): 117–46. doi.org/10.2307/4486062.

and proceed solely as an iterative digital publication.[24] The online *Stanford Dictionary of Philosophy* provides 'authoritative, rigorously accurate knowledge, at no cost to readers'.[25] Alongside arxiv.org (which serves academic physics and mathematics),[26] these nonprofit projects, hosted by institutional libraries,[27] may be instructive to national biographical dictionaries, whose future lies primarily online.

The transformation of information from analogue to digital formats alters how we access and consume it. Our ability to access copious information reifies the value of considered knowledge. Herbert Simon suggests 'a wealth of information creates a poverty of attention and a need to allocate that attention efficiently among the overabundance of information sources that might consume it'.[28] The increasing volume of information amassed within the digital revolution, reinforcing the value of reference works, has been transformative for national biographical dictionaries.

Attribution and Use

The mode of access of a reader can influence how the *DIB* is considered. Revision of the *DIB* has created a divergence between the static hard-copy entries and their amended online versions. We have found examples of scholars relying on an updated online *DIB* entry, yet referencing the

---

24 Alastair Jamieson, 'Oxford English Dictionary Will Not Be Printed Again', *Daily Telegraph*, 29 August 2010, accessed 7 September 2018, www.telegraph.co.uk/culture/books/booknews/7970391/Oxford-English-Dictionary-will-not-be-printed-again.html.
25 Nikhil Sonnad, 'The Philosopher's Stone: This Free Online Encyclopedia Has Achieved What Wikipedia Can Only Dream of', 21 September 2015, accessed 21 September 2015, qz.com/480741/this-free-online-encyclopedia-has-achieved-what-wikipedia-can-only-dream-of; Allen, Nodelman, and Zalta, 'The Stanford Encyclopedia of Philosophy', 210–28; Joint Conference on Digital Libraries, July 2002, accessed November 2016, plato.stanford.edu/pubs/jcdl2.pdf.
26 'arXiv.org', Cornell University Library, arxiv.org. 'PsyArXiv', Psychology's dedicated open-access digital archive, launched on 16 December 2016, promising 'to create free, open access to psychological science, even for papers that are ultimately published in journals that are only accessible to subscribers'. David Barner, Benjamin Brown, and Alex Holcombe, 'Introducing PsyArXiv: Psychology's Dedicated Open Access Digital Archive', 8 December 2016, accessed 11 August 2017, blog.psyarxiv.com/psyarxiv/2016/12/08/psyarxiv-press-release/.
27 Stanford University Library and Cornell University Library, respectively.
28 H. A. Simon, 'Designing Organizations for an Information-Rich World', in *Computers, Communication, and the Public Interest*, ed. Martin Greenberger (Baltimore, MD: The Johns Hopkins Press, 1971), 40–41.

2009 hard-copy volumes.[29] The static multivolume edition, organised alphabetically by surname, contrasts with the online platform, which can be interrogated in various ways. The well-understood form of discrete ownership, implicit in the former, contrasts with what is effectively a form of rental agreement governing access to the latter.

The *DIB* was extensively drawn upon in recent works spurred by the Irish 'decade of centenaries' (1912–22). Citation form varies, with the *DIB* sometimes cited *en masse,* when it is clear a small number of specific entries have been used—easily identifiable by the presence of memes exclusive to the *DIB* being reproduced, unattributed, in the text. Neither a monograph nor an edited collection, the *DIB* endures inconsistent citation. Either the entire multivolume work (by title and editors, and perhaps volume), or the discrete *DIB* entry (by title and author alone) are given. These two forms are in turn subject to a wide variety of construction and formatting, which undermines citation and impact analysis.[30]

Many scholarly sources can now be accessed through online platforms; for example, a printed journal article accessed as a PDF through JSTOR.[31] Should scholars cite the analogue form of a resource when it has been accessed digitally? Online material (for example, newsletters, blog posts, or online comments) is inherently difficult to reference. The *DIB* provides as much information as possible (title, author, date, and URL). The static representation of dynamic online resources will remain problematic.[32]

---

29 Evidenced by their usage of specific information only found in the revised online entry. This pattern is reinforced by Jonathan Blaney, 'The Problem of Citation in the Digital Humanities', conference presentation, 7 September 2012, accessed 9 August 2017, www.sheffield.ac.uk/polopoly_fs/1.209016!/file/JonathanBlaney.pptx. Blarney's case studies examining British History Online are available at 'British History Online', *TIDSR: Toolkit for the Impact of Digitised Scholarly Resources*, Oxford Internet Institute, University of Oxford, last modified 14 May 2011, microsites.oii.ox.ac.uk/tidsr/case-study/348/british-history-online.

30 'The footnote varies as widely in nature and content as any other complex scientific or technical practice … footnotes appear in enough forms to challenge any taxonomists ingenuity.' Anthony Grafton, *The Footnote: A Curious History* (Cambridge, MA: Harvard University Press, 1999), 11.

31 'JSTOR provides access to more than 10 million academic journal articles, books, and primary sources in 75 disciplines.' 'About JSTOR', JSTOR, accessed 14 August 2017, about.jstor.org/.

32 Uniform Resource Locator is a specific form of, and sometimes confused with, U(niform) R(esource) I(dentifier), akin to a web address. D(igital) O(bject) I(dentifiers) serve a similar function in controlled research environments, their operation governed by an ISO. For a useful discussion of persistent identifiers, see Ryan Moats, 'URN Syntax', memo, May 1997, accessed November 2018, www.rfc-editor.org/rfc/pdfrfc/rfc2141.txt.pdf; Emma Tonkin, 'Persistent Identifiers: Considering the Options', *Ariadne*, no. 56 (2008), accessed November 2018, www.ariadne.ac.uk/issue56/tonkin?lt%3B%2FA=&gt%3B=&lt%3B%2FP=.

In the United Kingdom, the Joint Information Services Committee, which supports higher education research, undertook a sequence of surveys in 2009 and 2010, assessing how 'British History Online' project resources were referenced. They found a bias towards academics encouraging students to cite hard-copy versions of sources, even when those sources have been accessed exclusively online, emphasising the absence of agreed citation standards across historical publications.[33] Lara Putnam notes that digitised material is more likely to be used and cited, and analogue material less so, referencing Tim Hitchcock's findings that, despite online access and keyword searching, ensuing bibliographic references are predominantly to hard-copy versions of texts.[34]

Despite efforts to archive selections of the internet, much online content will perish. What is captured may not be rendered at its original location or in its original form. Capturing Irish-related material is a key goal of the National Library of Ireland.[35] Such institutional efforts, alongside voluntary endeavours undertaken by the likes of the Internet Archive,[36] face significant technical difficulties and resource constraints. We can only hope that such noble efforts to capture resources integral to future biographical research are successful.

33   JISC, 'The Impact and Embedding of an Established Resource: British History Online as a Case Study', Jonathan Blaney and Peter Webster, Institute of Historical Research, School of Advanced Study, University of London, March 2011.

34   Putnam quotes Hitchcock saying that '[t]he vast majority of both journal articles and early modern and nineteenth-century printed sources are now accessed online and cherry-picked for relevant content via keyword searching. Yet references to these materials are still made to a hard copy on a library shelf, implying a process of immersive reading'. Lara Putnam, 'The Transnational and the Text-Searchable: Digitized Sources and the Shadows They Cast', *American Historical Review* 121, no. 2 (2016): 388. doi.org/10.1093/ahr/121.2.377.

35   Charlie Taylor, 'Ireland's Digital Content in Danger of Disappearing, Specialist Warns', *Irish Times*, 20 July 2017, accessed 25 July 2017, www.irishtimes.com/business/technology/ireland-s-digital-content-in-danger-of-disappearing-specialist-warns-1.3157792.

36   'Internet Archive is a non-profit library of millions of free books, movies, software, music, websites, and more.' Internet Archive, accessed 14 August 2017, archive.org/.

Ireland Online

The Bureau of Military History collections and the 1901 and 1911 census of Ireland, both recently digitised, have transformed biographical research in Ireland.[37] The digital edition of the *1641 Depositions*, over 8,000 witness accounts of a rebellion that year in Ireland, serves early modern scholarship.[38] Newspapers, periodicals, and serials continue to be digitised.[39] The digitisation of Irish civil and religious records (detailing births, marriages, and deaths) has made slower progress.[40] Recent work to harmonise digital humanities best practice, centred around standardising data formats, interoperable metadata, and linked data technologies, have been spurred by the impressive Digital Repository of Ireland.[41] The impact of these projects has been enormous, especially in designating familial relationships amongst and between *DIB* subjects.[42]

Websites such as Findagrave, Archiseek, and the Commonwealth War Graves Commission sit alongside a growing range of ephemera housed at the Internet Archive.[43] The latter's 'Wayback Machine' captures versions of deleted, defunct, or otherwise lost websites.[44] The ability to interrogate such digital resources was almost unimaginable during initial preparation of the *DIB* in the 1990s. Enthusiasts and voluntary bodies collect online obscure publications, newsletters, fanzines, sporting memorabilia, and

---

37 'Bureau of Military History', Military Archives and National Archives, www.bureauofmilitary history.ie/; 'Census of Ireland 1901/1911 and Census Fragments and Substitutes, 1821–51', National Archives of Ireland, www.census.nationalarchives.ie/. For a history of Irish Digital Humanities see James O'Sullivan, Órla Murphy, and Shawn Day, 'The Emergence of the Digital Humanities in Ireland', *Breac: A Digital Journal of Irish Studies*, 7 October 2015, accessed 9 August 2017, breac.nd. edu/articles/the-emergence-of-the-digital-humanities-in-ireland/.

38 '1641 Depositions', Trinity College Dublin, 1641.tcd.ie. A collaboration between Trinity College Dublin, the University of Aberdeen, the University of Cambridge, and IBM, the project launched in 2010.

39 See the work of the 'Newspaper and Periodical History Forum of Ireland', accessed November 2018, newspapersperiodicals.org/.

40 Various resources are collected at 'Irish Genealogy.ie', accessed November 2018, www.irish genealogy.ie/en/. A range of private sector and voluntary services aggregate abstracts of the same records.

41 'Digital Repository of Ireland', Royal Irish Academy, repository.dri.ie/; A. O'Carroll and S. Webb, *Digital Archiving in Ireland: National Survey of the Humanities and Social Sciences*, National University of Ireland Maynooth, 2012. doi.org/10.3318/DRI.2012.1. See also 'The Placenames Database of Ireland', accessed November 2018, www.logainm.ie/.

42 The *DIB* online is extensively cross-referenced, with over 60,000 cross-reference hyperlinks between over 10,000 articles.

43 'Find a Grave', secure.findagrave.com; 'Archiseek', archiseek.com; 'Commonwealth War Graves Commission', www.cwgc.org; 'Internet Archive', accessed November 2018, archive.org.

44 'Explore more than 304 billion web pages saved over time.' Wayback Machine, accessed 14 August 2017, archive.org/web/.

other ephemera. As artists, cultural figures, and 'celebrities' enter the *DIB* in greater numbers, we increasingly rely on such material to account for their lives.

Divergent methodological practices govern hard-copy and online publication. The scholarly and technical provenance of the digital resources we rely on must be considered. Veracity, explicitly addressed on the printed page through attribution and citation, collectively emerges from the content and structure of a hard-copy volume. A survey of historians' methodological approaches to digital humanities resources identified a widespread and limited understanding of the basic technical foundations of search functionality.[45]

Hard-copy access is mediated by indexes and page heads. Online access is digitally mediated, reliant on interwoven layers of hardware, software, and data. Technical implementation is as important as any scholarly methodology when presenting published or archival material online. Technical considerations subtly inflect methodological and editorial issues, cumulatively governing scholarly utility. Their interdependence is unfortunately widely ignored or misunderstood.

Scholarly Considerations

We assume that search results presented to us are neutral. They are in fact arbitrary, deliberately conditioned by algorithms. Their operation, opaque at best, influences user behaviour, tailoring the search results presented to us. It has been argued that 'we have replaced our guardians of information with algorithms that are dumb and that can be toyed with and manipulated'.[46] The perceived opportunity to search widely and deeply is illusory. The search results presented to us, lacking objectivity, are subjective abstractions.

---

45   Fred Gibbs and Trevor Owens, 'Building Better Digital Humanities Tools: Toward Broader Audiences and User-Centered Designs', *Digital Humanities Quarterly* 6, no. 2 (2012), accessed November 2018, digitalhumanities.org:8081/dhq/vol/6/2/000136/000136.html.
46   Carole Cadwalladr, 'Google "Must Review Its Search Rankings because of Rightwing Manipulation"', *Guardian*, 5 December 2016, accessed 11 August 2017, www.theguardian.com/technology/2016/dec/05/google-must-review-its-search-rankings-because-of-rightwing-manipulation.

We complacently search vast swathes of digital content without considering how digital resources are designed.[47] There is a widespread disinterest in how search functionality operates. Ethereal in nature, search functionality is construed as a technical issue to be discussed by geeks. Algorithms limit what users are presented with, determining how digital content is accessed and interrogated. The ubiquity of digital search functionality is unfortunately mirrored by widespread ignorance of its operation.

Janine Solberg notes how 'the digital tools and structures that increasingly support our research efforts have material and epistemological implications for how we discover, access and make sense of the past'.[48] Such considerations—assessing the provenance and accuracy of ensuing findings—have been internalised within traditional humanities scholarly practices. Humanities scholarship must similarly assess the work of the database creator, designer, and publisher.

Mitchell Whitelaw suggests 'search fails to match the ample abundance of our digital collections'.[49] A survey of the digital research practices of humanities scholars identified a prevalence of keyword searching, in contrast to how little used advanced search options were. The authors identify a paradoxical attitude of scholars towards searching: 'provenance and context are deemed key academic qualities, [yet] these do not appear to be common considerations in digital research practices'.[50]

D. Sculley and Bradley Pasanek urge scholars to examine and account for their own biases and assumptions. In the face of the 'superficial objectivity of computational methods', humanists must deploy and observe scholarly boundaries between computational results and their interpretation, analysis, and presentation.[51] The scholarly use of digital resources requires

47   Evgeny Morozov, *To Save Everything, Click Here: Technology, Solutionism, and the Urge to Fix Problems That Don't Exist* (London: Penguin UK, 2013).
48   Janine Solberg, 'Googling the Archive: Digital Tools and the Practice of History', *Advances in the History of Rhetoric* 15, no. 1 (2012): 54. doi.org/10.1080/15362426.2012.657052.
49   Mitchell Whitelaw, 'Generous Interfaces for Digital Cultural Collections', *Digital Humanities Quarterly* 9, no. 1 (2015): 3, accessed 2 November 2016, digitalhumanities.org:8081/dhq/vol/9/1/000205/000205.html#p3.
50   Max Kemman, Martijn Kleppe, and Stef Scagliola, 'Just Google It—Digital Research Practices of Humanities Scholars', *arXiv:1309.2434 [Cs]*, 10 September 2013, last revised 22 April 2014, 16, arxiv.org/abs/1309.2434.
51   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): 409. doi.org/10.1093/llc/fqn019.

due consideration of how they are created and published. The engineering and technical considerations that govern their use must be properly assessed.

Scholarly conventions require the standardisation of dates, name forms, and other named entities. Yet such issues are documented in only the highest quality digital humanities resources and databases. Digital technology is a means to an end. We should outline the parameters it operates within, ensure our objectives are met, and reject novelty for its own sake.

Algorithms, OCR and False Negatives

Erroneous indexing and metadata, alongside subjective algorithms and personally tailored search results, have a monumental impact on how we interrogate digital resources. Digital newspaper archives use optical character recognition (OCR) to scan hard-copy text. The impact of OCR technology from the 1980s was truly revolutionary, allowing masses of material to be digitised. When automated OCR generates a transcription error, left in situ or rectified by an automated error-correcting algorithm (accepting the error, or 'correcting' it—erroneously or otherwise), a new version of a text is produced. This new algorithmically generated intermediate text is inserted between the user and the underlying source material. This has profound implications for scholarly practice. Search algorithms and OCR technology together frame much of our biographical research.

For example, in researching an individual in the *Irish Times*, identical search parameters generated 19 results in the ProQuest database, compared with 378 results in the paper's own digital archive.[52] When queried, the ProQuest IT team could not explain this massive anomaly. This example reinforces the need to try out name variants and a variety of approaches when using digital archives. Certainly, the distinct onomastics of Irish names require variable search techniques. However, unless a researcher suspects search results are anomalous (here the researcher's intuition

---

52 The search term was 'Fred O'Donovan', the date range 1 January 1938 to 31 December 2010, and the search was undertaken on 22 June 2016 at both the *Irish Times* newspaper archive (www.irishtimes.com/search) and the ProQuest portal of the same (search.proquest.com/hnpirishtimes/). I am hugely grateful to my colleague Dr Linde Lunney for bringing this example to my attention.

emanated from decades of professional biographical research experience), such a false negative is highly likely to be accepted at face value. In this case, familiarity with the area being researched led to the conscious questioning of the results presented.

Each newspaper article in a database is a rendition of the original from which metadata (to facilitate indexing, search, and retrieval) is drawn under (at best) limited human supervision. Understandably, trade-offs between cost, efficiency, and accuracy impact large-scale digitisation projects. Scanning and indexing millions of analogue items is expensive. Yet technical and editorial provenance has profound implications for scholarly users. Digital databases are not pure renditions of their analogue content; errors are inevitable. Databases must be recognised for what they are: erratic and incomplete digital abstractions. David Berry has observed how 'little understood is the way in which the digital archives being created are deeply computational in structure and content … [c]omputational techniques are not merely an instrument wielded by traditional methods, rather they have profound effects on all aspects of the disciplines' using them.[53] Thomas Hughes observes that engineered systems are not mere technical or scientific apparitions, but the result of complex social, cultural, and political processes.[54]

Our subconscious reliance upon algorithmically delivered results subtly denudes our inquisitiveness. Algorithms embody the biases of their creators, delimiting the results presented to users. The return of false positives is obvious (incorrect answers to our queries). The return of false negatives (when a negative result indicates, incorrectly, the absence of something that is in fact present) remains unseen. Any faults in a book, such as imprecise or limited indexing, will become apparent after extended use. The technical structure of a digital resource or collection, and the metadata drawn upon by the search algorithm, are hidden from a user. Furthermore, as search functionality mediates our access to digital

---

53   David M. Berry, 'The Computational Turn: Thinking About the Digital Humanities', *Culture Machine* 12 (2011): 13, accessed November 2016, www.culturemachine.net/index.php/cm/article/view/440.

54   Thomas Hughes, *Human Built World: How to Think About Technology and Culture* (Chicago and London: University of Chicago Press, 2004), quoted in James Smithies, 'A View from IT', *Digital Humanities Quarterly* 5, no. 3 (2011), accessed November 2018, digitalhumanities.org:8081/dhq/vol/5/3/000107/000107.html.

resources, it precludes the 'serendipity of browsing'.[55] Chance occurrences during casual browsing have been the spark for several transnational lives and 'missing persons' added to the *DIB* since 2009.

Our past search queries and online behaviour are tracked, to personalise the search results presented to us. This creates an echo chamber.[56] Different users, using the exact search query, may return different results, whether they are sitting next to each other or on different continents. Such technology, regarded as valuable intellectual property, is explicitly hidden from view. We are presented with results the algorithm *thinks* we want, not necessarily *all* results for a specific search term.

Some have argued for 'algorithmic accountability' to audit the 'black boxes' that govern online search. A recent review of the impact of algorithms argued for their assessment across five principles: responsibility, explainability, accuracy, auditability, and fairness. We must consider algorithms as human creations and consciously address their inputs and outputs so as to mitigate their inevitable errors and biases.[57] Sue Halpern recently concluded 'there is a tendency to assume that data is neutral, that it does not reflect inherent biases … We need to recognize that the fallibility of human beings is written into the algorithms that humans write'.[58]

Putnam, discussing how full-text searching has become the new norm, elucidates how 'the new topography of information has systemic blind spots' and 'opens short cuts that enable ignorance as well as knowledge'.[59] Digital interrogation, for Putnam, subtly conditions what questions are 'efficiently answerable, and therefore worth asking'.[60] A grim circularity is sadly possible. Material that remains undigitised may be less visible to

---

55   'The serendipity of browsing … has yet to be successfully recreated in electronic form.' Steven Poole, '*You Could Look It Up* by Jack Lynch, Review—Search Engines Can't Do Everything', *Guardian*, 21 April 2016, accessed August 2016, www.theguardian.com/books/2016/apr/21/you-could-look-it-up-jack-lynch-review.

56   Carole Cadwalladr, 'Google, Democracy and the Truth About Internet Search', *Guardian*, 4 December 2016, accessed 9 August 2017, www.theguardian.com/technology/2016/dec/04/google-democracy-truth-internet-search-facebook.

57   Nicholas Diakopoulos and Sorelle Friedler, 'How to Hold Algorithms Accountable', *MIT Technology Review*, 17 November 2016, accessed 9 August 2017, www.technologyreview.com/s/602933/how-to-hold-algorithms-accountable/.

58   Sue Halpern, 'They Have, Right Now, Another You', *New York Review of Books*, 22 December 2016, accessed 9 August 2016, www.nybooks.com/articles/2016/12/22/they-have-right-now-another-you/.

59   Putnam, 'The Transnational and the Text-Searchable', 375.

60   Putnam, 'The Transnational and the Text-Searchable', 367.

scholars. Such material, as it cannot be searched digitally, may then be less likely to be consulted. Information that remains analogue is in danger of being ignored. We must consciously address the technical foundations of the digital resources we use.

Conclusion

Humanities research increasingly relies on digital technologies, mediated by the ubiquitous search box. Biographical research, into transnational lives and those at the margins of history, underserved in existing historiography, has benefited considerably. Yet our reliance on digital research, especially our dependence upon search algorithms, requires conscious consideration. Objective searching of vast digitised collections is somewhat illusory. Our trust in the search results presented to us (either positive or negative) is unwarranted.

Our scholarly inquisitiveness requires us to challenge the results presented to us by digital resources. In doing so, we must recognise how little we understand the technical and engineering principles that govern our digital research. That is not to sanctify the tangible, analogue outputs of traditional scholarship and publishing. Only by recognising deficiencies in our research practices can we address flawed assumptions that may undermine biographical scholarship. Traditional values—cautious scholarly assessment and considered editorial review—remain paramount as we face the digital future.