The Old Bailey Online at 20

David Hitchcock

Published online: 06 Aug 2024.

Submit your article to this journal ⍁

Article views: 102

View related articles ⍁

View Crossmark data ⍁

VIEWPOINT

The Old Bailey Online at 20

DAVID HITCHCOCK

Canterbury Christ Church University, UK

It is an interesting challenge to review a longstanding digital humanities project after 20 years in public operation. After such a duration its users will know whether or not it has been successful, and in this case, there is little doubt. Providing detailed and searchable online access since 2003, and in more or less constant operation since that time, today *The Old Bailey Proceedings Online, 1674–1913* (hereafter 'Old Bailey') is one of only a handful of digitised archive projects that might plausibly claim a broad public currency in the UK, not least because of recent (in)famous adaptations and misuses of its contents.[1] I would venture to guess that almost every upper-year university class on the history of crime in England across the UK and North America has engaged with the Old Bailey (certainly that is where I first learned of it, as a student in an honours class about criminal and legal history in Ottawa in 2006). For many historians of my generation and geographical focus, the Old Bailey and Early English Books Online were our introductions to the powers and possibilities of doing history digitally. I would not be surprised if sources from the Old Bailey are cited in a substantial percentage of publications both academic and popular that touch on crime, society, and plebian lifeways and lifecycles. In fact, a Zotero bibliography updated to 2017 before discontinuation lists 646 publications that directly cite the website, and I am sure many more publications could be added to it.[2] In some ways, the site is too good *not* to use, which affects historiographies. More on that later.

By what criteria should we measure the success of a digital humanities project designed to make broadly available a very extensive physical archive? Presumably, it is necessary to consider the longevity and stability of the resource, the fidelity with which it engages the real archive it 'facsimulates', the flexibility and utility of the tools it provides, the usability of its interface, the extent and availability of the collection to any given user, and how transparent the project's methods are. On all these scores it seems clear that the Old Bailey is a 'sector leader' and has been so from its inception. There is no paywall or institutional

limiter on the site's data or functionality. Each update of the project has worked to improve searchability and to add in tools and refinements, from 'fuzzy searching' to record linkage to visualisation to the addition of huge new collections, such as the 72 million words of material in the 'post-1834' proceedings added in updates over a decade ago. The site is relatively user-friendly, users do not need much training to deploy its advanced search features, and detailed tutorials are provided that can teach any user how to use the more advanced functions. When you have refined your search, individual results generally take you straight to a facsimile and transcription without undue fuss, and then suggest any connected materials for further consideration.

We should also contemplate the effects of the database more widely after so long in sustained operation: how might consistent text-searchable access to one major criminal archive have affected or even transformed a range of scholarships and fields, from law to literature to sociology to, of course, social and legal history? Finally, it is important to think about the team of scholars who—across 20 years—have designed, maintained, updated, and cared for the Old Bailey site, particularly its creators Professors Tim Hitchcock and Bob Shoemaker, and Dr Sharon Howard and Jamie McLaughlin. Some members of this team have worked with the site long enough to pen their own scholarly retrospectives covering a decade of engagement, and I hope they look back again now.[3] What is clear is that consistent institutional support matters enormously to the sustainability of large digital resources. I am quite sure the site's traffic and activities exert a significant draw on the servers at the Sheffield Digital Humanities Institute where it is 'housed', and the site requires a modest funding turnover to continue. Nevertheless, what we have in the Old Bailey is a vanishingly rare thing: a 'digitally alive' primary resource collection that is old enough to vote, which has been updated and maintained across many iterations of the worldwide web, open to all users, which works to *archive itself* rather than leave that important task to chance, and which tracks the manifold impacts it has made over time.[4] Few scholarly digitisation projects can claim a similar record of success.

'Version 9.0' has only recently been released, and for the remainder of this review I propose to explore from a user's perspective what this latest iteration of the site enables and arguably what it affects in my scholarly 'patch': early modern British social and cultural history.[5] First and foremost, the updated search function boasts excellent flexibility, and must set the bar for multi-criteria query construction and easy Boolean (AND, OR, NOT) and proximity ('x is found within n words of') searches. The search engine itself has clearly been geared towards students and scholars of history; we can set a query to run only inside particular record types (front matter, advertisements, legal records), and we can control by date range, defendant, victim, offence, location of crime, and more. Powerful new functions are also available, such as the ability to generate basic statistics from a set of results, all controlled and x/y indexed using the same data categories as the search queries used. Setting a date limit of before 1800, I can see within a few clicks that the keyword 'vagrant' returns twenty-five records, in which twenty-two men and six women appear before the court for various offences, mostly theft, although one woman, Mary Grout (not actually a vagrant

it seems), appears to have committed a crime against royalty (counterfeiting coinage) in 1798.[6] 'Vagabond' returns another thirty hits, 'beggar' another thirty on top of that. Within moments I can formulate not just a research dataset, but entire visualisations and even envision nascent projects to pursue.

The database now also integrates occupational data about victims of crime, a major step forward in fleshing out individual cases and wider social contexts. The revamped website hosts seven new 'historical background' pages including an important pair on the court's links to empire and on the presence of South and East Asian people in the record set. The updated African Diaspora page briefly outlines the 'continuous Black presence in London since at least the sixteenth century' and draws on up-to-date population estimates from Simon Newman's book *Freedom Seekers* which, as it is an open-access monograph, the summary page could consider linking directly through to (with similar practice followed elsewhere).[7] These pages collectively create a solid teaching and undergraduate research resource. Students can be directed to population history pages to unpack details about London's size and occupational distribution, and to pages on coinage and wages to understand how much money people might have made and how much they had to spend, all over a long duration. The information is up-to-date and so can double as efficient ways for students to cite key quantitative details about the city at least to 1913. The bottom of each page hosts an introductory reading list, though direct citations are omitted. New guides on how to use the site to 'do statistics' also feature, though technically the site itself adeptly 'does the statistics' for a user based on parameters they set.[8] However, this guide provides excellent and succinct explanations of different statistical visualisations (pie charts, bar charts, various graphs) and the relative strengths of each. In all cases, the supplementary guidance and summary pages seem clearly designed with classroom teaching in mind, across a solid range of age groups and levels, but primarily with History undergraduates in view.

It remains now to consider for a moment how the wide success of the Old Bailey over 20 years may have affected scholarship, and here I limit myself only to two arenas: digital 'headaches' or the unintended consequences of long-running projects, and a kind of 'scholarly bend' in which London continues to disproportionately feature in certain types of history-writing in no small part due to how rich its surviving primary source-base is, and how successfully resourced the digitisation efforts focused on it have been. 'London' appears in the title of at least 421 British history books about the period before 1900 published between 2014 and 2024.[9] It appears in far more book titles as an index subject. The historiography on the capital city is enormous and valuable, but to demonstrate the extent of the focus on it, consider that 'Newcastle' appears only eleven times —and Bristol only fifty-eight times—using the same search parameters.

I am aided in these considerations by a 2015 special issue of *Law, Crime, and History* on 'our criminal past', where Sharon Howard made several important contributions. First of all, Howard was already worried then that the Old Bailey Online had bent research attention disproportionately towards London and towards higher criminal courts more or less regardless of the larger scholarly context. While I think the wider academic historical profession will naturally

'correct' for these imbalances over time, I do think this worry applies to scholarship that is not focused on crime in history, but which turns to crimes *in* history for various forms of resource, evidence, and inspiration (but then might not consider the limitations of the 'Proceedings' primary sources on which the database is built). The collective obligations of regional, social, and cultural differences in the patterns and predispositions of crime and criminality tend to broaden out our criminal historiography in a way that protects against too lingering a focus on a single metropolis, though of course an excellent array of books on crime in London have long benefitted from, arguably even relied on, the website.[10] But the 'bend' is visible in other disciplines too. Consider the Old Bailey 'corpus', that is to say the 127 million or so machine language words that it currently contains, which has been used quite frequently by linguists and 'distant reading' projects owing to its size, chronological breadth (useful in tracking linguistic shifts), and provenance, but it is worth asking if those projects have taken seriously the important limitations and considerations imposed by the formatting, genre, as well as geographical, survival, and selection biases of the record set itself as they 'mine' through its contents.[11]

Howard also wrote that 'vast swathes of British crime and punishment archives —especially the local and provincial—remain entirely untouched by digitisation', which is still sadly true today, and noted that the Old Bailey was 'in the right place at the right time' when digitisation money flowed freely from a New Labour government and relatively few projects were in competition for it.[12] We know now that seemingly utopian digital moment is long gone, while questions about the longevity and expensive maintenance of large digital artefacts created in previous decades remain with us. I count it a great opportunity missed, for instance, that English 'Quarter Sessions' papers are almost never found digitised in a machine-readable format, despite having been frequently transcribed and published in editions by record societies on account of their immensely varied and interesting contents. Howard also mentions two important follow-on digital projects connected to the Old Bailey, *London Lives* (which remains rather like an echo of an earlier version of the Old Bailey Online, last updated in 2018) and *Connected Histories* (last updated in 2019). What happens if one of those breaks? Here we confront the spectre of digital dependency, and not just of one database on another, but of entire architectures like hyperlinks in footnotes (DOIs can only ever partially address this issue), digital accessioning method variations, and so on. How *long* will these techniques of citation remain meaningful? I ask because the footnote has been going strong for about half a millennium in printed works and is, if anything, even more relevant as an information architecture today. Like the project team who have so successfully updated the Old Bailey and prepared it for its next decade of digital life, we need to think seriously about the long-term health of our big digital history projects, *and* about the ways we refer our future readers back to them.

We also need to think about a new danger: automated content mining for the 'training' of large language models. At no fault whatsoever of the project, but as a byproduct of the very radical openness at the heart of it, the Old Bailey Online might soon feature—indeed I would be unsurprised to see aspects of it 'sold'—as

a training tool for so-called 'lawbots', legal-focused LLM iterations that are currently very trendy at large UK and US solicitor firms.[13] Will its custodians for the foreseeable future resist this fiduciary temptation? On environmental considerations alone the implications are worrying, given that the average energy and water usage of a single LLM query is twenty-five times greater than a single Google search. What ethical considerations might now apply to the use and misuse of this vast corpus of names, occupations, crimes, judgements, deliberations, verdicts, and precedents?

What it is worth saying in conclusion is that the Old Bailey Online is a unique and remarkable resource; a huge collective contribution to what we know, and what we can know, about London's ordinary, extraordinary, and criminal past. As it enters its third decade of existence, the website boasts a thoroughly updated backend functionality, a tidy user interface, and robust new tools to aid researchers. Its creators and curators should be justly proud of their accomplishments while remaining alive to emerging dangers of abuse and misuse of this popular resource.

Notes

1   'Voices from the Old Bailey', *BBC Radio 4* (2010–2014), https://www.bbc.co.uk/programmes/b012stwb; Alison Flood, 'Naomi Wolf Accused of Confusing Child Abuse with Gay Persecution in Outrages', *The Guardian* (8 February 2021), https://www.theguardian.com/books/2021/feb/08/naomi-wolf-accused-of-confusing-child-abuse-with-gay-persecution-in-outrages.

2   'Old Bailey Proceedings Online: Citations Bibliography', https://www.zotero.org/groups/31512/old_bailey_proceedings_online_citations_bibliography/library [accessed 29 June 2024].

3   Sharon Howard, 'Bloody Code: Reflecting on a Decade of the Old Bailey Online and the Digital Futures of Our Criminal Past', *Law, Crime, and History*, 1 (2015), 12–24.

4   'Digital Projects Using Old Bailey Online Data', *The Old Bailey Proceedings Online, 1674–1913*, https://www.oldbaileyonline.org/about/projects#past [accessed 29 June 2024].

5   'What's New (Autumn 2023)', *Old Bailey Proceedings*, https://www.oldbaileyonline.org/about/whats-new [accessed 29 June 2024].

6   Cross-tabulation of offence category (rows) against defendant gender (columns), *Old Bailey Proceedings*, https://www.oldbaileyonline.org/search/statistical?initviz=table-cross&month_lte=1&rows=defendant_gender&series=

offence_category&text=vagrant&year_lte=1800#results [accessed 1 May 2024].

7   'The African Diaspora', *Old Bailey Proceedings*, https://www.oldbaileyonline.org/about/black [accessed 1 May 2024].

8   'Doing Statistics', *Old Bailey Proceedings*, https://www.oldbaileyonline.org/about/doingstatistics [accessed 29 June 2024].

9   This number derived from an advanced search using the Bibliography of British and Irish History with the following limiters: title contains 'London'; publication dates, 2014–2024; type, book; dates, 1500–1900.

10 Most notably Hitchcock and Shoemaker's own *London Lives*, which experimented with digital footnoting; but also books like Jensen's *Vagabonds*, Gattrell's *Conspiracy on Cato Street*, and Conley's *Women Who Killed*, which relies enormously on the database, so much so the date range in its title owes to the Old Bailey's coverage (1674–1913). See: Tim Hitchcock and Robert Shoemaker, *London Lives: Poverty, Crime and the Making of a Modern City, 1690–1800* (Cambridge: Cambridge University Press, 2015); Oskar Jensen, *Vagabonds: Life on the Streets of Nineteenth-Century London* (London: Duckworth Books, 2022); Vic Gatrell, *Conspiracy on Cato Street: A Tale of Liberty and Revolution in Regency London*

(Cambridge: Cambridge University Press, 2022); and Carolyn A. Conley, *Debauched, Desperate, Deranged: Women Who Killed, London 1674– 1913* (Oxford: Oxford University Press, 2020).

11   Indicatively, see Section 2, 'Background', in: Peter J. Grund, 'Disgusting, Obscene and Aggravating Language: Speech Descriptors and the Sociopragmatic Evaluation of Speech in the Old Bailey Corpus', *English Language and Linguistics*, 27.3 (2023), 519–520.

12   Howard, 'Bloody Code', 13.

13   See: 'Risk Outlook Report: The Use of Artificial Intelligence in the Legal Market', *Solicitors Regulation Authority* (2023), https://www.sra.org.uk/sra/research-publications/artificial-intelligence-legal-market/ [accessed 29 June 2024].