'all writing is in fact cut ups':[1] the UK Web Archive and Beat literature

Rona Cran, University College London

'the historical sense involves a perception, not only of the pastness of the past, but of its presence'[2]

1.0   Introduction

In the 1950s a group of American writers including William Burroughs, Allen Ginsberg, Jack Kerouac and Gregory Corso came to prominence as part of a cultural phenomenon known as the Beat Generation. Simultaneously documenting and embodying Beat culture, Burroughs, Ginsberg, Kerouac and Corso were held up, both by themselves and by others, as promoters and producers of explicit, innovative texts, users of a range of illegal drugs, rejectors of materialism and tradition, and experimenters in alternative sexualities and religions. Perhaps the most enduring aspect of their literary work is the aleatory cut-up technique, pioneered by Burroughs. A form of collage, the cut-up technique involves slicing or folding up pages of text, and then rearranging them in a random order, in a quest 'to form new combinations of word and image'.[3]

Focusing on the period between 1996 and 2013,[4] during which the transition to the internet age can broadly be said to have taken place, I planned to attempt to use web archives to analyse both public and academic receptions to Beat literature, in the UK, in the late 20th and early 21st centuries. Although the idea of the Beats and Beat literature remains popular in the cultural imagination, serious scholarship in the field of Beat studies is often marginalised in the context of contemporary academia. Such scholarship (although usually excellent) often struggles to find a home in print, and is located primarily on the internet, on public websites such as *Reality Studio* (www.realitystudio.org) and the *European Beat Studies Network* (www.ebsn.eu). Because of this, the online interaction of Beat scholars with Beat enthusiasts (non-scholars) and other writers is extensive, as is the resulting output, whether it takes the form of literary articles, online debate, uploaded conference proceedings or interested questions.

However, much of this material, while internationally available, is hosted either by US or mainland European websites. I therefore planned to use the UK Web Archive to locate key (though possibly now obsolete) UK-based websites or webpages including Beat-related blogs, online magazines, fansites, forums, academic websites and comments sections, before analysing the information retrieved thematically, geographically and chronologically. I established three key research questions:

- How has reception to the Beats developed during the period in question, and has the more democratic nature of internet discussion changed scholarly and non-scholarly responses?

- Can web archives be used as a valid research tool for academic and non-academic work pertaining to Beat literature and wider literary studies?

- How can scholars and members of the public engage more efficiently with online and archived Beat-related academia, and ensure that their future work is archived accessibly?

In hindsight, my research aims for this case study were too ambitious, given the timeframe, the size of the archive, and the prototypical nature of the SHINE interface. When planning my case study, I envisioned extensive analysis, via the live web and the UK Web Archive, of both public and academic receptions to Beat literature in the late 20th and early 21st centuries. My background is in literary and archival research and criticism, and, not really knowing all that much about the UK Web Archive, I imagined that this would be the easy part – reading about and analysing literature about literature is, after all, what I do. My research proposal also suggested, more tentatively, the creation of a working index of search terms which produce informative and relevant results when searching for Beat-related material on the UK Web Archive, as well as a Beat-related sub-archive or corpus, on which other researchers would be able to build. Perhaps the most surprising result of my case study was that the former proved to be almost impossible within the scope of the project, while the latter – more technical – suggestions are the ones that have proved more fruitful, and which will be more valuable, ultimately, in helping me to achieve the initial aims of my project. This is testament to the ongoing development of the SHINE interface, which currently offers two different search options (one basic and one advanced), and a number of filtering mechanisms. When I first began my case study in early 2014, for instance, simply searching for the term 'Beat literature' returned innumerable useless results, such as obsolete Amazon pages, but the development of an advanced search function and filtering options has helped to hide these where appropriate, enabling me, to some extent, to exclude irrelevant entries and get rid of surplus results. The capacity to save searches and create corpora goes and will continue to go a long way toward enabling researchers to carry out, over time, just the sort of extensive and ambitious projects that I initially had in mind, as well as demystifying the UK Web Archive and therefore making it enduringly useful to researchers regardless of their levels of familiarity with complex IT systems.

2.0   Note on methodology

This write-up of my case study deliberately prioritises methodological reflections over any detailed and substantive answers to my research questions. There are two main reasons for this decision. The first is relatively simple, and will, I hope, be borne out below: I believe that a more generalised reflection on the value and workings of the UK Web Archive will be of greater consequence to future researchers as and when they come to use it. The second reason is linked to this, but is more complex and relates to a key limitation of the dataset, which researchers are advised to bear in mind when using the archive. My hope, when starting this project, was to locate equivalent websites to *Reality Studio* (currently the most dynamic and comprehensive site dedicated to William Burroughs) or *The Allen Ginsberg Project* (www.allenginsberg.org), in order to map and analyse UK-based scholarship on and interest in the Beats. However, the material I discovered in the Archive was far more fragmented and disparate, and as such would require a very different methodology to the one I used. I did not conclude that Beat scholarship or interest does not or did not exist in the UK – rather that either it had not found its way onto the internet in the late 20th and early 21st centuries, or UK-based scholars and readers contributed to or used other domains, such as .com, .org or .eu. In the light of this, my substantive conclusions are, at this point, relatively small, as well as being more speculative than evidence-based. I now hope that I can extend my case study beyond the remit of

the Big UK Domain Data for the Arts and Humanities project and use other internet archives (including the Internet Archive (www.archive.org) and the UK Data Archive (http://www.data-archive.ac.uk) to carry out further research in this area. Nevertheless, given the highly innovative nature of this particular project, and the need to develop a theoretical and methodological framework for using the UK Web Archive, I hope that my methodological observations will be useful in shedding light on the process of searching and data collection and in opening up web archives as an important resource for Arts and Humanities researchers worldwide.

2.1   Foraging and sense-making: mapping the archive

My initial forays into the UK Web Archive led me to refine my research questions to the following:

1. Can web archives be used as a valid research tool for academic work pertaining to Beat literature and wider literary studies?

2. How can scholars engage more efficiently with archived Beat-related academia?

3. How useful is the UK Web Archive for scholars trying to find alternative (non-academic) responses to Beat literature, and how easy are such responses to find?

4. How has reception to the Beats developed during the period, and has the more democratic nature of internet discussion changed scholarly and non-scholarly responses?

Broadly speaking, in answer to the first two questions, I found that the UK Web Archive has great validity and enormous potential as a research tool for literary researchers, particularly as we move further forward into the internet age and an increasing amount of relevant and important material is recorded in a digital format only. There is a liberating sense, when working within the archive, of exploring both the past and the future, simultaneously – of entering uncharted territory while also rediscovering forgotten artefacts (although possibly this is an aspect that may be lost as the search interface develops). In order for scholars and members of the wider public to interact more efficiently with the UK Web Archive, at this stage it simply needs to be used – the territory needs to be mapped, in other words. This is something that researchers can do, I have found, by experimenting with search criteria and search functions, by familiarising themselves with any relevant constraints, and by carrying out a process of foraging and sense-making, not just within the context of their own individual research plans but also within the archive itself.

The limitations of the archive are not immediately obvious to the untrained eye, and where possible, researchers should endeavour to work alongside the developers of the archive, in order to more fully understand what is happening when they search for data. This will enable scholars to be alert to the potential subversion of expectations and to take into account any relevant constraints or parameters, such the ways in which descriptions are optimised, the significance of crawl dates, the likelihood that a specific capture may have failed to encapsulate a whole website or even a whole page,[5] and the fact that .uk does not represent the entirety of UK webspace. For instance, my initial research proposal favoured the use of sentiment analysis when searching for data, in order to assess any cultural developments or trends, but this was quickly discovered to produce misleading and inaccurate results, and so was discounted. As it stands, the UK Web Archive is an increasingly important resource for Arts and Humanities researchers but an awareness of the key tenets of its

functionality is essential if one is to embark on any kind of sustained project that relies predominantly on the data it contains.

With regard to my third question, it is evident that scholars who use the UK Web Archive to mine untapped or forgotten or otherwise unavailable resources will have a clear advantage over their peers, providing they take into account the limitations noted above. They will also have the potential to add a new dimension to the study of Beat literature, by, for instance, charting its popularity (or otherwise) over a given period, mapping any fluctuations in readership or response, by creating, for instance, a small corpus which could be scaled up using an algorithm or analysis tool. While the material I discovered using the dataset was deeply fragmented, a different approach or methodology, whereby specific websites are identified for study in advance, coupled with an extended project time-frame and a more profound knowledge of the SHINE interface, seems likely to prove fruitful. It may seem an obvious point, but there is an enormous amount of material available in the archive (approximately 65 terabytes and many billions of words) – enough for many different researchers to consider many different aspects of Beat literature, on its own terms as well as in relation to the Web Archive. Searching blind, as I did, comprehensive responses to Beat literature within the dataset are, simply put, not easy to find, but approaching the archive with some familiarity with the terrain, whether in terms of context or practical training (or both), will help to ameliorate this.

My final question remains very much a work in progress, and I will discuss it in more detail below. I initially found the sheer volume of relatively un-curated material in the archive to be something of a setback. I considered narrowing my focus from the Beats in general to William Burroughs in particular, concentrating less on the admittedly rather nebulous 'contemporary imagination', upon which my research proposal was based, than on more specific public responses to the 50th anniversary of Burroughs's most famous work, *Naked Lunch*, which took place in 2009. However, I concluded that this particular area of study in fact warranted its own project, and so I continued with my original, albeit slightly revised, case study. While I have compiled and am continuing to compile a range of scholarly and non-scholarly responses to Beat literature using the Web Archive, analysis of this vast and disparate collection of sources in a methodical and coherent manner will take more time. As it stands, it seems more relevant and appropriate to report, albeit speculatively, on what I see as the larger picture regarding Beat literature and the UK Web Archive, and of course internet archives in general. As I will explain below, it became evident that in many ways the UK Web Archive, particularly in its current prototypical state, is an ideal mechanism through which to approach the study of the Beats, and, as such, my approaches to the archive and what I view as its essential and inherent messiness, are inextricably linked to my final research question.

2.2   Approaches to the archive: chance meetings vs. the 'Google mindset'

The sheer volume of material available in the Web Archive is staggering. To further complicate matters, researchers' familiarity with advanced live-web search engines – or what fellow researcher Richard Deswarte refers to as 'the Google mindset' – falsifies expectations regarding what the SHINE search interface can offer (although, as I will discuss, this can be liberating as well as restrictive). While a vast amount of material, in itself, is every researcher's dream, in this case dealing with hundreds of thousands of results served to put the ambitions of my case study into perspective. Searching blind, using limited search terms, often proved painstaking, time-consuming and unrewarding. As Saskia Huc-Heffer discovered, in her case study, it is generally more effective to

approach the archive with some understanding of the terrain, such as prior awareness of a specific host or website.

As I have noted above, this was something I lacked – and had, indeed, hoped to find. However, I discovered that by taking a deliberately unsystematic approach to the archive – by treating it as something akin to a vast bundle of unsorted papers rather than, say, Google – I was able to confront it with my own perspectives, rather than having my perspectives neatly reflected back at me, as the live web is wont to do. In the absence of Google's high-tech (and highly expensive) algorithms, I was able to decide for myself which results were and were not relevant to my case study, heightening its intellectual integrity by using processes of reasoning and selection which were unique to me. This process led me to question American commentator Leon Wieseltier's notion, articulated in his article 'What big data will never explain', that 'in the riot of words and numbers in which we live so smartly and so articulately, … we are renouncing some of the primary human experiences'.[6] This is certainly an idea with which many Arts and Humanities researchers will identify, but I found that by approaching the UK Web Archive as a physical library or manuscript repository I was fully able to preserve, as a researcher, what Surrealist André Breton called 'the intoxicating atmosphere of *chance*'.[7] Peter Webster observes that the 'arbitrariness' of making one's own decisions within the archive rather than relying on a closed algorithm leads to 'a kind of collage in the Web Archive'.[8] In my experience this was certainly the case. Collage – a key practice within Beat literature – prizes creativity, resourcefulness and serendipity, and, to quote Breton again, 'has the marvellous faculty of attaining two widely separate realities without departing from the realm of our experience, of bringing them together and drawing a spark from their contact'.[9]

I found, therefore, that the lack of structure and curation in the archive was, for my purposes, useful. By going in blind, as it were, I was able to make my own guesses and draw my own conclusions, not knowing what I would find. Unlike the live web, the archive did not try to explain itself or to second-guess me, to finish my sentences, to assume certain search preferences, or to rank search results based on perceived popularity. As a result, I was able to discover more unusual or obscure blogs, personal webpages and fansites than I would have been able to on the live web. I had a sense of entering into a new discipline, and, as a scholar relatively new to the field of Digital Humanities, I enjoyed being able to work freely, unencumbered by complex theories or previous research practices.

In relation to this, it soon became apparent that such an archive is, in some senses, the ideal system through which to study the Beats. William Burroughs, Jack Kerouac, Allen Ginsberg and Gregory Corso were writers who published their own and each other's work, who wrote entire novels on single scrolls of paper, in Kerouac's case, or who had their work typed up by several different people and delivered to the publisher in haphazard sections over a number of weeks, in Burroughs's. Burroughs's novels, in particular, read – if they can be said to read at all – like the uncontrolled spewings of an ailing machine. Just as Burroughs's texts imply that in some alternate universe the narrative may once have been fully functional, but now trails only the ghost of familiarity, so the UK Web Archive resembles the live web but is fundamentally different from it. The UK Web Archive is to the internet what Burroughs's cut-up texts are to traditional novels – related yet alien, and requiring the deconditioning of innate responses in order for understanding to occur.

The Beats' *modus operandi* was spontaneous, performative and collegial, very much like the internet, and consequently much like the UK Web Archive. Messiness, for the Beats, was a way of life. Inconsistency was not something they worried about. Missing items, uncorrected errors, differing versions of what should be the same thing, and works in progress are among the most interesting and significant elements of the Beats' work, and to find these aspects mirrored on the Web Archive was revelatory, if challenging. Burroughs, Ginsberg, Kerouac and Corso, among others in the Beat Generation, treasured notions of fragmentation, ellipsis and inherent unknowability, all of which, for me at least, are positive aspects of the Web Archive in its present form. In relation to this, it is important to note that like many archives, a web archive is continually expanding, and as with a physical archive, not everything can be preserved. It is helpful, therefore, when using the UK Web Archive, to keep in mind the similarities between online and offline archives (and to forget about the all-encompassing nature of Google), in order to actively draw attention to what has been omitted, and why. Just as missing diary entries or letters may prove crucial in the context of a physical archive, so un-crawled or un-captured websites may also prove to be significant within the context of the Web Archive. As the Archive develops, questions over who (or what) decides what is or is not relevant will inevitably be raised, and the scholarly integrity of the Archive scrutinised. It is important, therefore, not to ignore the unavoidable gaps in the dataset but to address them head on.

The Beats also, of course, fully embraced the 20th-century's technological developments and would have relished the scope and ambition of the development of the archive. In fact, there are striking similarities between the fundamental nature of the internet (and subsequently the Web Archive), and the ways in which the Beats, in their pre-internet era, disseminated their creative material. Both the Beats and the internet in general are relatively informal and highly collaborative, often performative, with varying and often indefinable audiences, international readerships, various formats, fluid, experimental and unedited texts, evolving or devolving material, and numerous different versions of the same or similar things. From the websites – predominantly blogs, fansites, literary magazines and personal websites – that I have examined so far, it is clear that people's responses to the Beats mirror the work of the Beats themselves, and that consequently the internet is a model vehicle for such responses, enabling constant revision, debate and collaboration, and also enhancing the extent of interaction and exchange. For instance, the archived pages of *The Richmond Review,* a now obsolete literary magazine established in October 1995 and billed as 'the UK's first literary magazine to be published exclusively on the World Wide Web', include links to other Beat-related websites (mostly in the US), detailed, hypertextual biographies of Beat figures, tributes to the Beats, reposted articles or texts by Beat writers, and even poetry (particularly haikus) inspired by one or more of the Beats. Many of the links are broken or no longer exist, and many of the pages were seemingly not captured or crawled, but the sense of what the Beat-enthusiasts at *The Richmond Review* were trying to do – informatively, collaboratively and creatively – is very much in evidence.

3.0   Conclusion

Far from being devoid of 'primary human experiences', the UK Web Archive seems to me to be replete with them. As Brewster Kahle – the founder and creator of the San Francisco-based Internet Archive – asserts, 'the Net is a people's medium: the good, the bad and the ugly. The interesting, the picayune and the profane. It's all there'.[10] That the work and ideas of individuals who engaged, formally and informally, with Beat literature and thought during the late 20th and early 21st

centuries has been digitally preserved for future generations is significant in itself. While the UK Web Archive will no doubt go on to become a highly comprehensive and well-oiled machine, the incomplete and fragmented version with which I have been working for much of my case study has added a fascinating dimension to my study of the Beats – one which William Burroughs, who spent much of his life cutting texts apart and sticking them back together in a different order, would particularly have appreciated, as, of course, do I. For me, the relationship between the Beats and the UK Web Archive is only just beginning.

---

[1] William Burroughs, 'The cut-up method of Brion Gysin', in *A Casebook on the Beat*, ed. Thomas Parkinson, (New York, 1961), pp. 105–6.

[2] T. S. Eliot, 'Tradition and the individual talent', in *The Sacred Wood* (1920; London, 1997), p. 40.

[3] Burroughs, quoted in Barry Miles, *William Burroughs: El Hombre Invisible* (London, 1993), p. 119.

[4] The Big UK Domain Data for the Arts and Humanities project works with the dataset derived from the UK domain crawl from 1996 to 2013 (when legal deposit legislation was extended to cover digital materials).

[5] Websites over 10MB in size may not be captured in their entirety.

[6] Leon Wieseltier, 'What big data will never explain', *New Republic*, 26 March 2013.

[7] André Breton, *Mad Love* (*L'Amour Fou*), trans. Mary Ann Caws (Lincoln, Nebr., 1987), p. 25.

[8] Peter Webster, 'Method in the web archive for the arts and humanities: a conference report', 20 January 2015 (http://peterwebster.me/2015/01/20/method-in-the-web-archive-for-the-arts-and-humanities-a-conference-report/ [accessed 19 June 2015].

[9] André Breton, 'Preface to the Max Ernst Exhibition, May 1920', in Max Ernst, *Beyond Painting* (New York, 1948), p. 177.

[10] Brewster Kahle, in Lee Dembart, 'the end user / A voice for the consumer: Go Wayback', *The New York Times*, 4 March 2002 (http://www.nytimes.com/2002/03/04/business/worldbusiness/04iht-itend04_ed3_.html [accessed 19 June 2015]).