1.

‘The past went that-a-way’: editing in the rearview mirror?

Andrew Prescott

One of Marshall McLuhan’s most celebrated metaphors was what he called the rearview effect. McLuhan pointed out how our reaction to new technologies is shaped by our previous experience. We do not immediately grasp the potential of new technologies but interpret them in the light of what we know. In McLuhan’s words:

> When faced with a totally new situation, we tend always to attach ourselves to the objects, to the flavour of the most recent past. We look at the present through a rear-view mirror. We march backwards into the future (McLuhan, Fiore and Agel 1967, 74–5).

Moreover, McLuhan suggested, we rely on driving by the rearview mirror because the view it offers may be more familiar and comforting than the alarming prospects visible through the windscreen. To quote McLuhan again:

> Ordinary human instinct causes people to recoil from these environments and to rely on the rear-view mirror as a kind of repeat or *ricorso* of the preceding environment, thus ensuring total disorientation at all times. It is not that there is anything wrong with the old environment, but it simply will not serve as a navigational guide to the new one (McLuhan and Parker 1969, xxiii).

Although McLuhan's overall analysis is not entirely convincing (the early explorations of radio, for example, show strong experimental instincts and were not always shaped by past experience), McLuhan's metaphor is compelling. The rearview effect can be seen at many points in the history of technology. When Lewis Cubitt was asked to design an early railway terminus at King's Cross, he took as his model the Czar's Riding Academy in Moscow. In naming parts of aeroplanes, we looked back at the terminology used in ships, so that some of this nautical terminology is also used in spacecraft. The history of text technologies provides many examples of the rearview effect. The earliest books printed with movable type frequently imitated the appearance of manuscripts. Similarly, early photographers such as Julia Margaret Cameron used photography to create scenes which were like historical paintings. In McLuhan's words, 'We impose the form of the old on the content of the new' (McLuhan, Fiore and Agel 1967, 86).

The rearview effect pervades our approach to digital technologies. Computers retain a qwerty keyboard designed for mechanical typewriters, complete with a carriage return key (although we increasingly refer to it as an 'enter' key). We use metaphors from the world of printed books and manuscripts to describe different forms of handling information in computers – files, libraries, archives. The rearview mirror is not only apparent in the way in which computers are designed and built, but also in the way we use them. The design of spreadsheets is rooted in the structure of ledger books and other forms of accounting stationery. A simple relational database like Microsoft Access looks back to card indexes and punch card sorting. Images are kept in albums. Are we really using the power of computers in completely new ways, or is McLuhan correct in his observation that 'Our official culture is striving to force the new media to do the work of the old' (McLuhan, Fiore and Agel 1967, 81)? Anybody who has had to struggle with the kind of corporate systems used in institutions such as universities might be inclined to agree with McLuhan.

Digital editions were one of the early success stories of the World Wide Web. Imaginative digital editions quickly appeared of canonical

works ranging from *Beowulf* (ebeowulf.uky.edu) to John Foxe's *Book of Martyrs* (www.dhi.ac.uk/foxe/), together with digital archives of the works of figures such as William Blake (blakearchive.org), Dante Gabriel Rossetti (http://www.rossettiarchive.org/) and Walt Whitman (whitmanarchive.org). Commentators such as Jerome McGann, Peter Robinson and Hans Walter Gabler extolled the benefits of digital editions in providing multifaceted views of texts, particularly through the use of hyperlinks which would enable easy access to the primary manuscript, printed or other materials on which the edition depended (McGann 1991; Gabler 2010; Robinson 2010). The possibilities offered by automated collation and search also seemed to offer potential for improved methods of tracing the genealogy of a text, although this has largely proved a chimera. Nevertheless, it seemed that digital potentialities would foster a renaissance in editing as a mainstream scholarly activity. Peter Robinson observed in 2010 that 'It is a truth universally acknowledged that all papers on scholarly editing these days must contain the word "revolution"' (Robinson 2010, 57). In describing how this revolution had now reached a quiet phase, Robinson inevitably reached for the comparison with Gutenberg: 'Gutenberg's bible was a shot heard around the world; we are still living through the transformation of our culture which followed' (Robinson 2010, 57) .

The reference to Gutenberg occurs in much of the literature on digital scholarly editing, and it alerts us to the possibility that McLuhan's rearview mirror may be in play here. The history of editing is inextricably bound up with print. While humanist scholars had already made great progress in the critical analysis of texts prior to Gutenberg, it was the arrival of print that spawned the development of the edition. Patrick Sahle offers us a broad definition of an edition with his formulation that 'A scholarly edition is the critical representation of historic documents' (Sahle 2017, 23). The need for such a critical representation is driven by the requirement to reproduce authoritative texts in different media – manuscripts in print, print or manuscripts in digital form and so on. There is an assumption that an editor will seek to correct errors in the manuscript or other text which is being reproduced. This was succinctly summed up by John Mitchell Kemble in his 1833 edition of *Beowulf*:

> A modern edition, made by a person really conversant with the language which he illustrates, will in all probability be much more like the original than the MS copy, which, even in the earliest times, was made by an ignorant or indolent transcriber. But while he makes the necessary corrections, no man is justified in withholding the original readings: for although the laws of a language, ascertained by wide and careful examination of all the cognate tongues, of the hidden springs and ground-principles upon which they rest in common, are like the laws of the Medes and Persians and alter not, yet the very errors of the old writer are valuable... (Kemble 1833, xxiv)

From this formulation by Kemble, we can see how all the various forms of editorial practice and the disagreements about editorial procedures sprang up. As soon as a corrected form of the text claims superior authority, and the need to show the evidence for that is accepted, all the various forms of editing, from diplomatic editing through to the need for simplified teaching editions, inevitably flow. The shape of these editions and the conventions used to express the status of the text are driven by the need to present the text in printed form. Much of our conception of the edition springs from that comforting image in the rearview mirror of the opulent, stately and beautifully crafted printed scholarly editions of the nineteenth and twentieth centuries.

How are our assumptions about the future of digital editing shaped by the rearview mirror? How far are we ignoring the problems coming towards us that are visible in the windscreen? As we start to confront the issues involved in making accessible radically new types of primary information sources, will the editorial procedures of the nineteenth and twentieth centuries provide any guide? I suspect such precedents will be of limited value. Indeed, I wish to suggest that the very concept of an edition is a backward looking one, an artefact of the rearview mirror. While the need to present authoritative and accessible literary, historical and other texts will, I imagine, remain a constant need, increasingly we will be dealing with born-digital data, so that the idea of what a 'critical representation' might

constitute will need fundamentally rethinking. The role currently fulfilled by editions might increasingly be fulfilled by visualisations or APIs for metadata. Digital forensics may play a key role. The editor (if such there is) might have very little to do with the actual email or social media texts, but be much more concerned with the interfaces and status of the text. Is this the death of the edition? No – it is a development of our existing editorial and critical skills to deal with completely new types of material. But many key features thought to be characteristic of the scholarly edition will need to be rethought and re-imagined as we grapple with new types of born-digital environments.

There was in the 1990s an assumption that the inherent advantages of digital editions meant that they would become generally preferred for scholarly purposes, but this has not proved to be the case. For many scholarly editors, the gold standard remains the reassuring sense of permanence offered by print editions produced by major scholarly publishers such as Oxford University Press. The AHRC-funded *Editing Robert Burns for the 21st Century* project at the University of Glasgow has as its focus a multivolume print edition of *The Works of Robert Burns*, published by Oxford University Press. The digital component comprises a website with performances of songs and readings from Burns's works (burnsc21.glasgow.ac.uk). *The New Oxford Shakespeare*, produced under the leadership of Gary Taylor, John Jowett, Terri Bourus and Gabriel Egan, also adopts a hybrid approach.[1] For scholars, a two-volume printed *Complete Works of Shakespeare* with original spelling, press variants and so on is being published, while students and more general readership are offered a separate one-volume *Complete Works* with modern spelling and punctuation. An authorship companion aimed at scholarly users is also being produced. All four projects are available online via the *Oxford Scholarly Editions Online* platform. The role of commercial publishing platforms such as *Oxford Scholarly Editions Online* further complicates matters. Digital scholarly editing

1 https://global.oup.com/academic/category/arts-and-humanities/literature/shakespeare/new-oxford-shakespeare/?cc=us&lang=en&.

specialists usually consider that print editions made available online are not true digital editions (Sahle 2017: 27–33), yet the widespread library access to commercial packages such as *Oxford Scholarly Editions* means that this form of digital edition will be extensively used by students and researchers.

The way in which print practices have been carried over to digital editions is particularly apparent in the editing of historical documents. The preparation of summaries known as calendars to provide access to the voluminous contents of administrative records has a venerable history stretching back to at least the seventeenth century (Ramsay 1960; Johnson 1960; Knighton 2007). When programmes for the large-scale publication of public records were set in hand in Great Britain in the nineteenth century, priority was given to the publication of calendars of chancery records. However, the preparation and publication of such summaries was expensive, both in manpower and in printing costs. By the time of the publication in 1977 of *Editing Records for Publication* by R. F. Hunnisett, a senior archivist in the Public Record Office, the drive to reduce printing costs had become paramount, and Hunnisett recommended that no post-1300 records should ever be printed in full because of the cost of printing (Hunnisett 1977, 14–16). Rereading Hunnisett's manual today is like visiting a lost world. The discussion is dominated by typographic conventions and ways to make printing cheaper and more efficient.

The high cost of the publication of calendars and the fact that they diverted resources from managing and making available current archival accessions meant that the production of record calendars had hugely declined by 1990. Geoffrey Elton loudly criticised the way in which calendars encouraged historians to rely on short and misleading abstracts so that they never looked at the archives (Elton 1969, 90–2; Cantwell 2000, 53–7). It might be thought that the arrival of the World Wide Web might have provided an opportunity to rethink methods of publishing historical archives. Manfred Thaller in his 1992 Duderstadt project set out not only to digitise the entire archives of a small town in Germany but also to explore the nature

of the continuum between digitisation and the edition (Thaller 2017, 44–5). However, the rearview mirror effect kicked in and, far from exploring new forms of access, historians seized on the World Wide Web as a means of reviving the moribund project of producing calendars. Projects such as *Mapping the Medieval Countryside* (inquisitionspostmortem.ac.uk) and *The Gascon Rolls Project 1317–1467* (www.gasconrolls.org) are a revival of the Victorian series of calendar publications, even to the extent of following Hunnisett's recommendations for editorial procedure, despite the fact that many of these suggestions are designed to reduce printing costs.

Contemporary government records do not look like medieval inquisitions, and editorial procedures designed to cope with the output of medieval chanceries will be of little value in making available government documents dating from the twenty-first century. The primary sources to which historians researching the twenty-first century will require access will be born-digital and they will be vast in scale. We can get a hint of their scale from the email archives of US Presidents. Correspondence has been a fundamental primary source of historians since the Renaissance, and printed editions of rulers and politicians have been at the heart of much historical research. When I started work at the British Library in 1979, I worked on the papers of the Duke of Marlborough and an indispensable aid to my work, consulted daily, was Henry Snyder's immaculate three volume edition of the correspondence between the First Duke of Marlborough and the Lord Treasurer, Lord Godolphin (Snyder 1975). For the later eighteenth century, it was possible for Arthur Aspinall to single-handedly produce compendious editions of the correspondence of George III and George IV (Aspinall, 1938; Aspinall 1963a; Aspinall 1963b), although the discovery of much additional material in the Royal Archives prompted the launch of The Georgian Papers (georgianpapers.com), a digital edition of this correspondence, by King's College London and the Royal Collection Trust. By the late nineteenth century, the expansion of information had become evident. The papers of William Gladstone in the British Library comprise approximately 160,000 documents, bound in 762 large volumes. Nevertheless, this is still a comparatively manageable

material and a small group of scholars might see Gladstone's correspondence as a large-scale, but manageable, project.

Contrast Gladstone's papers with the email archive of President George W. Bush. Email messages sent and received by each member of the White House staff during Bush's presidency are stored in the Electronic Records Archive of the US National Archives and form part of the George W. Bush Presidential Library. The system contains over 200 million email messages, The electronic records for Bush's Presidency amount to over 80 terabytes (Winters and Prescott 2019, 397). Massive though it is, the Bush archive is dwarfed by the official Presidential records of the Obama administration. 95 per cent of the records from the Obama administration are born-digital. There are approximately 1.5 billion pages of such born-digital records, including emails, PDFs, images and social media. The remainder of the Barack Obama Presidential Library comprises roughly 30 million pages of paper documents and 30,000 physical artefacts (www.obama.org/obama-archives/).

It is unlikely that anyone will easily be able to produce anything like a traditional edition of the presidential records of either George W. Bush or Barack Obama. The material is simply too vast. Moreover, a printed representation of these digital archives would lose a great deal of information. One of the most important elements of email is the address bar, which can be used to analyse who corresponded with whom, who was copied into particular emails and how emails were forwarded. Analysing the information in the address bar is only feasible if the digital record is used. The kind of printed representation that Snyder produced of Marlborough's correspondence or Aspinall for George III and George IV is neither practicable nor desirable for email archives like those of Presidents Bush and Obama. In accessing email archives, future historians will need to focus on metadata rather than the text of individual messages. The use of metadata by agencies like the UK's Government Communications Headquarters (GCHQ) and the National Security Agency (NSA) in the US to identify potential terrorist activity perhaps points to the sort of methods historians may have to use

in interrogating email archives (Winters and Prescott 2019, 397). In analysing the networks and other features revealed by emails, visualisations of, for example, Social Network Analysis will be important. It is likely that future editions of political correspondence will be visual representations of metadata rather than the stately volumes of a Snyder or Aspinall.

Email archives may seem intimidating enough, but they are straightforward compared to the problems which will be posed as born-digital government records become available. The range of born-digital archives currently accessible to researchers is comparatively limited and highly controlled, but we can get a good idea of the scale and difficulty of the problems that future researchers will encounter from leaks of sensitive government data such as the two tranches of logs documenting American military action in Afghanistan and Iraq released by Wikileaks in 2010, the American defence and security files leaked by Edward Snowden and the 11.5 million documents known as the Panama Papers, taken from a Panamanian law firm and detailing financial and client information for over 200,000 offshore entities (Assange et al. 2015; Bernstein 2019). These are precisely the sort of documents with which future historians writing the history of the wars of the early twenty-first century or reconstructing financial power structures will have to grapple.

While newspapers were quickly able to find sensational plums among this leaked material, the questions of how to represent the structure of such large-scale data and enable information easily to be retrieved are problematic. Julian Assange was urged to produce a printed edition from Wikileaks material, but the scale of the material and the difficulty of representing its interconnections made him hesitate. In introducing the volume which was finally produced, Assange emphasised how the printed edition was not really suitable for such material:

> Wikileaks has published 2,325,961 diplomatic cables and other US State Department records, comprising some two billion words. This stupendous and seemingly insurmountable body

> of internal state literature, which if printed would amount to some 30,000 volumes, represents something new. Like the State Department, it cannot be grasped without breaking it open and considering its parts. But to randomly pick up isolated diplomatic records that intersect with known entities and disputes, as some daily newspapers have done, is to miss 'the empire' for its cables (Assange et al. 2015, 1–2).

The Afghan war logs released by Wikileaks comprised 91,000 military records, while the Iraqi files consisted of 391,000 records. These were initially loaded into Excel, but the spreadsheet automatically truncated the import of the records after 66,000 records. Eventually a visualisation was produced (using as a template an interactive guide to the Glastonbury music festival) which allowed the attempts of the US Army to deal with improvised explosive devices in Afghanistan to be reconstructed day by day and year by year. For the first time, accurate death tolls of these military actions could be produced (Winters and Prescott 2019, 391–3). This initial visualisation shows a way forward, but of course the data can be analysed in many other ways. Geographers have used the Wikileaks data to map major insurgent clusters, to show how different types of attack occurred in different terrains, and to trace the intensity and violence of the conflict (O'Loughlin, Witmer, Linke and Thorwardson 2010).

Given the large quantities of data involved, machine learning and artificial intelligence approaches potentially have a great deal to offer. Successful experiments have been made with the use of self-organising maps to analyse the diplomatic cables released by Wikileaks. This methods uses machine learning to generate topic maps of large collections of born-digital data. Self-organising maps give a good overview of the overall concerns of the US state department and embassies in the early twenty-first century, with particular emphasis on, for example, the nuclear programmes of Iran and North Korea and the Russian-Georgian War of 2008 (Mayer and Rauber 2011). Social network analysis is also likely to figure prominently in approaches to born-digital records, and has been used very successfully with the Panama Papers. A social network analysis of the

Panama Papers has proved instructive in identifying patterns of the network structure of some offshore entities that are untypical and may help identify entities engaged in dubious business activities (Kejriwal and Dang 2020).

The maps, graphs and visualisations produced by analyses of born-digital records such as the Wikileaks material or the Panama Papers may arguably anticipate the type of edition of born-digital primary materials that historians of the future may need, but doubts might be felt as to whether the idea of an edition is at all helpful in this case. If the key feature of an edition is the representation of a text in another medium, are such visualisations of born-digital records a comparable representation to, say, a print edition of a manuscript text? It might be felt that visualisations form different functions and have a different scope from traditional editions. To produce such digital analyses, what is required is not so much an edition but rather clean, consistent data of known provenance and authority (something that, of course, inherently does not apply to Wikileaks material). Insofar as the precedents of printed editions are helpful here, it is in the importance of ensuring that the data is reliable and trusted and that its provenance can be traced. Another striking contrast between the requirements of born-digital analysis and traditional editions is the importance of automated tools in dealing with born-digital whereas in traditional editions it is the human intervention of the editor which is critical.

Many traditional forms of editing historical documents are not applicable to the types of born-digital materials on which historians will rely in the future. It may seem that this will be less of an issue with the literary texts more generally associated with discussions of editorial practice, but born-digital materials are already starting to appear in the literary archive and are also challenging conceptions of the edition. This material may not be on the same scale as the White House email archives or the Wikileaks diplomatic cables, but it is often more complex in structure and perhaps more directly challenges assumptions about editorial practice.

For example, the Scottish novelist Irvine Welsh is a prolific user of Twitter, having made over 94,000 tweets since joining Twitter in February 2012. He has at the time of writing over 370,000 followers.[2] Irvine Welsh's Twitter feed is interesting because Welsh writes the bulk of the tweets himself and describes his everyday life rather than engaging in commercial promotion. A moment's glance at Welsh's Twitter feed shows that it is potentially a very useful source for those interested in his life and work. Welsh's Twitter feed is a rich store of Welsh's humour and idiom, as on 14 December 2015, when Welsh tweeted:

> 'You're fuckin deid Welsh.' There. Just gave myself death threat to highlight issue of online abuse. That'll make them take notice. Or not.

To which @Calamity_Payne replied under the hashtag #GotYourBack 'I've reported it pal.'

There have been a number of academic studies on the relationship between football and literature in Welsh's novels (May 2016), and football figures prominently among Welsh's tweets, as for example in this thread published at 7.10pm on 13 October 2022 using Twitter for iPhone:

> If a team you support plays against a team who has 20 times more finance, it's pretty much given that your boys will not come out top. It's basic economics and it dictates our lives in the neoliberal order. If the team you support wins against a twenty times more impoverished...

2 Following its acquisition by Elon Musk in October 2022, Twitter was relaunched as 'X' in July 2023. Irvine Welsh's opinion of the relaunch is evident from his post on 24 July 2023: 'Some wide fucker of a designer had Muskie's keks down with this back-of-fag-packet work'. Welsh was reported to be leaving Twitter for Mastodon in November 2022 (*Glasgow Times*, 7 November 2022), but has maintained an 'X' account. His current profile reads: Typist. Woke cunt. Failed macrodoser. instagram: irvine.Welsh mastodon: @IrvineWelsh@mastodon.scot blueskies: @irvinewelsh.bsky.social.

> ...side, you would have to be a fuckwit to see this as a sign of your moral superiority. You were simply born in a bigger city, or worse, you're a shallow, glory hunting twat who only supports such a team to bolster your own inadequate self...
>
> ...and your own manifest failure to achieve anything in life. So you live by proxy through people who not only don't care whether you live or die, but worse, don't even know you exist in the first place. Outwith a few hundred extra ST quids in the accounts or the TV subs dosh.
>
> So enjoy football, whatever team you support, enjoy the banter, enjoy ripping the pish, but don't be a delusional cunt genuinely believing in your own moral superiority. This only advertises you as a total fucking loser.

Imagining what a critical edition of such a thread would look like poses a number of problems. I have retained here the division into tweets, indicated as in the tweets by ellipses, but the piece is written as connected prose. Should it be shown as a thread or as continuous prose? This tweet prompted lively responses from Welsh's followers. Do we include these in any edition? At one point in the exchanges, Welsh states that the tweet was meant as a message of support for Dundee United, which is clearly relevant information. Do we include just this response by Welsh, or provide wider contextual information?

Even more problematic is how an edition of the Twitter feed of an author like Welsh deals with the issue of metadata. A tweet is more than just text. Each tweet contains 150 data points, describing for example time, place, twitter client and device used, and account details. This information is potentially valuable for biographical and other purposes. In some cases, it may be vital for determining authorship. For example, it has been suggested that tweets by Donald Trump on an Android phone were made by Trump himself, but that tweets on his account from an iPhone were made by his staff (Robinson 2016). This claim has not been borne out by stylistic analysis (Clarke and Grieve 2019), but it indicates that a bare minimum in an edition of tweets should be device information. If

Twitter metadata is to be fully represented, can this ever be done in anything resembling a traditional literary edition? It is surprising that more attention has not been given to the complex editorial issues raised by Twitter. Again, it seems more practicable to work directly with Twitter downloads rather than any intermediary, but Twitter's increasing restrictions on third-party access to its data makes this difficult. A few of Irvine Welsh's tweets from 2014 to 2016 are included in the UK Web Archive, but these are not searchable, only a handful of tweets were harvested, and the profile has become garbled. It seems inevitable that literary scholars will seek to gather together the Twitter activities of authors like Welsh, Salman Rushdie, Margaret Atwood or Bret Easton Ellis – all active on Twitter – but it is not clear how a Twitter edition will function. And the problems are not restricted to Twitter. Irvine Welsh is active on Instagram, which poses another set of issues, particularly because of its pictorial content.

The letter has been a staple of literary scholarship and a major focus of traditional editing. Within a very short period of time, the literary letter has been replaced by the email. While the scale of emails beginning to appear in literary archives is much smaller than the millions generated by the Bush and Obama presidencies, nevertheless collections of emails included in the papers of authors pose challenging issues. For example, email archives are likely to include a great deal of sensitive personal information such as social security numbers or bank details. Trying to remove this before the emails are deposited in a library is very time-consuming and usually not completely successful. In order to ensure the authenticity of an email, access to metadata is often required. The threaded nature of many email conversations is difficult to represent in a form that enables users easily to follow the exchanges. The email archive of the poet Wendy Cope acquired by the British Library in 2011 comprised some 25,000 emails (Schneider et al. 2019; McKean 2020). The emails arrived in the library as a legacy PST file on a USB flash drive. In order to ensure all the available metadata was preserved, a forensic quality ingest was made into the library's eMSS system. In a sense, this forensic record of Cope's email archive may be

regarded as analogous to the physical manuscript of a traditional letter. Originally, the plan was to make the emails available in reading rooms in a PDF-A format, but the limitations of PDFs in exploring emails quickly became apparent. Much greater success was found when the ePadd platform developed by Stanford University and also used on the email archive of Ian McEwan at the Harry Ransom Center in the University of Texas was used. The tools provided by ePadd were more specifically designed for interrogating emails and would have greatly expedited the sensitivity review. It might be felt that the ePadd version of the Wendy Cope or Ian McEwan archives can be regarded as an edition – a representation of the original ingest which is more accessible for readers – but there are still major issues in, for example, the way ePadd searches attachments and difficulty in accessing technical metadata when required. Again, it is not entirely clear that the idea of an edition is a helpful metaphor in coming to terms with the problems poised by email archives. Thinking of an edition encourages us to imagine a fixed final representation, whereas with an email archive, the key consideration is establishing a workflow which preserves the integrity of the original archive but facilitates outputs which will meet the needs of both scholars and general readers.

If social media and emails pose problems enough, then the difficulties of the old Amstrad discs, floppy disks, CD-ROMs, DVDs, flash drives and hard drives that libraries and archives are increasingly accumulating are overwhelming. Projects such as the Salman Rushdie project at Emory University, which recovered word processing files from damaged and redundant Apple computers and made them available via an emulator, have shown what can be done (Farr and Waugh 2020), but the resources required are considerable and projects like the Rushdie project remain a rarity. More typical are stories of obstacles and difficulties in processing and making available born-digital materials. A recent survey by Lise Jaillant of access to born-digital archives in major British repositories paints a gloomy and sometimes alarming picture of born-digital records being acquired without access workflows being available, anxiety about the formidable legal and personal data sensitivity issues, and

shortages of resources and know-how. If born-digital archives cannot be made available in reading rooms, then the prospect of editions based on the born-digital seems very remote (Jaillant 2022).

The Wendy Cope archive acquired by the British Library included, as well as the email archive, 76 floppy disks of two types, 89.3 MB and 11.2 GB, saved on a USB flash drive. A workflow was developed to create a forensic copy of this material and then to generate PDF-A files for reading room access which were given the reference Add. MS. 89108. However, only a relatively small amount of the born-digital material in the Wendy Cope archive was processed and it is not currently included in the online catalogue (Pledge and Dickens 2017). The British Library acquired the archive of the writer Will Self in 2016. This included not only 541 files of diaries, correspondence, photographs, drafts, proofs and even Post-it Notes but also the contents of Self's computer hard drive comprising 100,000 emails and also (in the words of the blog entry announcing the acquisition of Self's archive) 'a wealth of electronic manuscript drafts and approximately 100,000 emails along with a huge number of other files yet to be mined and identified (including downloads of his i-Tunes, which offer an intriguing line of investigation for future users of the archive' (Foss 2017). While the manuscript component of Self's archive has been catalogued and made available with commendable speed and efficiency and are now under the overall reference code of Additional MS. 89203, it is not clear when and how the much larger born-digital elements will be catalogued and made available.

Since the arrival of the World Wide Web, the focus in digital scholarly editing has been on the creation of digital representations of works that first appeared in manuscript and print. The discussion has been chiefly about the advantages and disadvantages of presenting editions in a digital medium rather than in print. But the more pressing challenge is how we make born digital materials – email archives, government records, social media, word processing files – available for research and scholarship. In order to do this, we

require the development of workflows, which will address complex issues like sensitivity of personal information but at the same time enable the power of the metadata carried in born-digital files to be exploited. Given the huge scale of born-digital information, it is unlikely that this can be done by human intervention alone – some use of AI is inevitable.

In this sense, the idea of the edition has been something of a distraction, and it can be seen as an artefact of the rearview mirror. While editions still perform a function in providing trusted and rigorous representations of manuscript and printed texts, they offer little direct guidance on how to address the issues of access to born-digital information. It is the born-digital which increasingly fills the windscreen while we have been focusing on the rearview mirror. However, while the born-digital workflows on which libraries and archives will increasingly rely may bear little resemblance to the traditional edition, there is one area where they share key values, namely the importance of ensuring that information is grounded in the best quality data whose provenance is assured and whose structure and history can be investigated and tested.

References

Aspinall, A. 1938. *The Letters of George IV 1812–1830*. 3 vols. Cambridge University Press.

Aspinall, A. 1963a. *The Correspondence of George, Prince of Wales, 1770–1812*. 8 vols. Cassell.

Aspinall, A. 1963b. *The Later Correspondence of George III*. 5 vols. Cambridge University Press.

Assange, J. et al. 2015. *The Wikileaks Files*. Verso.

Bernstein, J. 2019. *The Laundromat: Inside the Panama Papers Investigation of Illicit Money Networks and the Global Elite*. Penguin.

Cantwell, J. D. 2000. *The Public Record Office 1959–1969*. Public Record Office.

Clarke, I. and Clarke, J. 2019. 'Stylistic Variation on the Donald Trump Twitter Account: a Linguistic Analysis of Tweets Posted between 2009 and 2018.' *PLoS ONE* 14 (9).

Elton, G. R. 1969. *The Practice of History*. Fontana.

Farr, E. and Waugh, D. 2020. 'Salman Rushdie Archive.' BitCurator Consortium. https://bitcuratorconsortium.org/salman-rushdie-archive/.

Foss, R. 2016. 'Will Self's Archive Acquired by the British Library.' *English and Drama* (blog), December 21. https://blogs.bl.uk/english-and-drama/2016/12/will-selfs-archive-acquired-by-the-british-library.html.

Gabler, H. W. 2010. 'Theorizing the Digital Scholarly Edition.' *Literature Compass* 7 (2): 43–56.

Hunnisett, R. F. 1977. *Editing Records for Publication*. Archives and the User 4. British Records Association.

Jaillant, L. 2022. 'How Can We Make Born-Digital and Digitised Archives More Accessible? Identifying Obstacles and Solutions.' *Archival Science* 22: 417–36.

Johnson, H. C. 1960. 'The Publication of English Records: the Public Record Office.' *Archives* 4 (24): 214–19.

Kejriwal, M. and Dang, A. 2020. 'Structural Studies of the Global Networks Exposed in the Panama Papers.' *Applied Network Science* 5 (63).

Kemble, J. M. 1833. *The Anglo-Saxon Poems of Beowulf, the Travellers Song and the Battle of Finnesburh*. William Pickering.

Knighton, C. S. 2007. 'The Calendars and their Editors, 1856–2006.' *State Papers Online 1509–1714*. Cengage Learning EMEA Ltd.

May, A. 2016. 'The Relationship between Football and Literature in the Novels of Irvine Welsh.' *Soccer and Society* 19 (7): 924–43.

Mayer, R. and Rauber, A. 2011. 'On Wires and Cables: Content Analysis of WikiLeaks Using Self-Organising Maps.' In *Advances in Self-Organizing Maps: WSOM 2011. Lecture Notes in Computer Science* 6731, edited by Laaksonen, J. and Honkela, T. Springer.

McGann, J. 1991. 'The Rationale of Hypertext.' In *Electronic Text: Investigations in Method and Theory*, edited by Sutherland, K. Clarendon Press.

McKean, C. 2020. 'Email Archives at the British Library: The Papers of Wendy Cope.' Presentation at Digital Preservation Coalition conference. www.dpconline.org/docs/miscellaneous/events/2020-events/2261-callum-mckean-wendycopeemailarchive.

McLuhan, M., Fiore, Q. and Agel, J. 1967. *The Medium is the Massage*. Allen Lane.

McLuhan, M. and Parker, H. 1969. *Through the Vanishing Point: Space in Poetry Painting*. Harper & Row.

O'Loughlin, J., Witmer, F., Linke, A. and Thorwardson, N. 2010. 'Peering into the Fog of War: The Geography of the WikiLeaks Afghanistan War Logs, 2004–2009.' *Eurasian Geography and Economics* 51 (4): 472–95.

Pledge, J. and Dickens, E. 2017. 'Process and Progress: Working with Born-Digital material in the Wendy Cope Archive at the British Library.' *Archives and Manuscripts* 46 (1): 59–69.

Ramsay, G. D. 1960. 'The Publication of English Records: Some Reflections on Mr Mullins's *Texts and Calendars*.' *Archives* 4 (23): 138–48.

Robinson D. 2016. 'Text Analysis of Trump's Tweets Confirms He Writes only the (Angrier) Android Half.' Variance Explained. varianceexplained.org/r/trump-tweets/.

Robinson, P. 2010. 'Editing Without Walls.' *Literature Compass* 7 (2): 57–61.

Sahle, P. 2017. 'What is a Scholarly Digital Edition?' In *Digital Scholarly Editing: Theories and Practices*, edited by Driscoll, M. and Pierazzo, E. Open Book Publishers.

Schneider, J., Adams, C., DeBauche, S., Echols, R., McKean, C., Moran, J. and Waugh, D. 2019. 'Appraising, Processing, and Providing Access to Email in Contemporary Literary Archives.' *Archives and Manuscripts* 47 (3): 305–26.

Snyder, H. L. 1975. *The Marlborough-Godolphin Correspondence*. 3 vols. Oxford University Press.

Thaller, M. 2017. 'Between the Chairs: An Interdisciplinary Career.' *Historical Social Research Supplement* 29: 7–109.

Winters, J. and Prescott. A. 'Negotiating the Born-Digital: A Problem of Search.' *Archives and Manuscripts* 47 (3): 391–403.