Digital News, Digitised News

Eldridge II, Scott

Digital News, Digitized News: Exploring online newspapers, and newspapers online
Scott A. Eldridge II

Timeline/Key dates for newspapers online
1991 – Universal Web Standards (http, html, URL, etc.) agreed to, start of 'world wide web'
1994 – Launch of web browsers Mosaic and Netscape, making 'the web' accessible to more users
1994 – Launch of the 'Electronic Telegraph', the first UK newspaper site
1995 – Launch of guardian.co.uk and 'Go2' online supplement
1997 – Shuttering of 'Shift-Control', increased news content on guardian.co.uk
1999 – Guardian.co.uk hits 1 million visitors
2006 – Guardian.co.uk goes 'digital first' with some news content

Timeline/Key dates for digital archives
1992 – British Library begins digitizing Burney collection
2004 – partnership with Gale Cengage and JISC to digitize 2 million pages of 19th century British newspapers
2013 – National Library of Wales makes digitized newspapers (pre-1910) fully available online
2013 – UK extends deposit rights to include British websites, allowing BL to develop web content archive.

Introduction
Much has been said of the digital turn within journalism, meant to describe the moment when news media began to take advantage of emerging online opportunities at the end of the 20th century. With the spread of the World Wide Web first in the United States and then in Europe in the 1990s, newspapers started to approach web platforms and their increased functionality and began to produce news not only for paper, but also for the screen. As they did, the idea of a 'digital turn' signalled more than an expansion of publication opportunities. It started a shift in the ways we think about news and media, informed both by what was now possible and what was expected. For newspapers and scholars alike, this introduced a complex set of dynamics that make this a particularly interesting moment in newspapers' histories. It is also a far less straightforward history than might be assumed, and newspapers who had developed their commercial approaches in the 20th century making a viable 'one-to-many' news product now stepped into an environment where open access, non-commercial experimentation, and interaction were commonplace. To make sense of unique approaches to sharing content now possible online, effort needs to be paid to exploring tensions between these media cultures, including digital cultures and newspapers' own occupational cultures to situate newspapers' transitions within these broader developments.

To do so, this chapter explores *The Guardian's* efforts to build an online presence as an exemplar of this transition and its tensions. Archives of its websites and print pages, and

accounts of those involved in developing its online spaces, offer insight into how newspapers' online identities first emerged in this new media space. From these we can also see where these opportunities were fraught with uncertainty. The discussion draws on the earliest archives of guardian.co.uk sites from 1995-1998 which, as media artefacts, reflect a range of decisions made by *The Guardian*. From there, this chapter points to what the digital turn meant for archives and research libraries transitioning from paper towards digitized archives and archives of digital content, and where lessons learned from newspapers' own transitions have had an effect on the digitization of archives, and archives of digital news. While distinct, each discussion highlights the opportunity and uncertainty that have accompanied the move towards a digital age.

Approaching newspapers' web histories

As a case study of newspapers embracing digital opportunities, *The Guardian*'s endeavour to go online demonstrates challenges at a conceptual level for considering newspaper websites, and at a practical level in trying to approach these for analysis. This chapter focuses first on archived webpages of *The Guardian*'s sites from the Internet Archive at archive.org using serendipitous discovery approaches for web and print archives (Foster and Ford, 2003; Toms, 2000). This follows links and archived sub-pages to explore content which query-based research alone would not uncover. While this approach is productive, it is not perfect. The archived sites offer an accessible and, in many cases, functional picture of these early sites, but as with any web archive there are broken links and occasionally missing pages which limit a fuller picture (Brügger, 2010a: 7). By also considering accounts of those involved in archives of *The Guardian*'s print pages, as well as other public accounts, these gaps are supplemented. These also help understand the socio-cultural aspects of newspapers' digital transitions. To do so, analysis bears three questions posed by Niels Brügger (2010b: 31) for considering web archives: "Why was there something and not just nothing? Why was there not something else? And why did it change?", while also weighing the relationship between "sender, medium, text, receiver and context" (Ibid.: 32).

While the internet itself was in use going further back in the 20th century (primarily across government and research centres, and for some news media for data transfers), this discussion focuses is on newspapers going 'online', meaning when they were publishing unique content on 'the web'. This draws attention to the mid-1990s as the underlying infrastructure of the internet was matched with a more user-friendly interface. This begins in August 1991, when familiar web standards rolled out including the http (hypertext transfer protocol), html (hypertext markup language), and URL (Uniform Resource Locators) standards (Brügger 2010a) . These (more or less) introduced the world to the 'World Wide Web', further made possible with the launch of browsers like Mosaic in 1993 and Netscape Navigator in 1994, bringing text and images together on screens for 'everyday' users. While it was 1998 before the web reached most of the world (Curran 2012: 35), the early 1990s and the release of more capable technologies and software brought 'the web' to Europe.

Though the introduction of the 'World Wide web' as a publicly-accessible space in the United Kingdom came in the early 1990s, few British news titles had websites then and those that existed were rudimentary (*Economist* 2012). It was not until the mid-to-late 1990s when news sites started establishing websites, first with *The Daily Telegraph* in 1994, followed by *The Guardian* in 1995, and the BBC in 1997, among others (Allan 2006). The *Guardian*'s first site at guardian.co.uk was the home of its 'New Media Lab', and in November that year Go2.guardian.co.uk launched as an extension of *The Guardian's* computer, science and technology print supplement, somewhat confusingly titled 'Guardian Online'. A year later, guardian.co.uk started to feature a wider range of coverage including news and sports content, with references to the site in the printed paper (*Guardian* 2010; guardian.co.uk 1996).

The Guardian's transition may have been late, then, but this should be considered with respect to the extent to which the Britain public went online. In its pages, *The Guardian* reported on 26 March 1995 that across Britain, only 3.5 million people had, at one time or another, accessed the internet – amounting to 8 percent of the population (McKie 1995). Though this was a faster uptake "than previously realised" (Ibid.), it shows we cannot speak of early newspapers sites as able to reach 'all', or even 'most', of a British populace. Instead, they were a relatively niche offering for a relatively niche audience, further specified in terms of geographies (55 percent of those in the South-east of England knew of the internet then, compared to 31 percent in Wales) and limited in use. Among the 'complex tasks' users engaged with online, downloading software, shopping, and researching products all ranked. Navigating news sites did not register.

Rather than dissuade us from considering early newspaper websites as important, their being accessed by a relatively small number of readers shapes how we look at such sites as places where newspapers were making sense of a new medium beyond the gaze of their wider audiences. In these early years, the web was an experiment and at guardian.co.uk, the platform is badged "a jumping off point for *Guardian* web projects" (guardian.co.uk 1995). This branding that gels with accounts of those behind the sites as well. Ian Katz (2009) refers to the first two years (1995-1997) of the guardian.co.uk website as a "sandbox" for New Media Lab projects. For *The Guardian*, this early period saw it present itself as an 'online newspaper', referring to something "conceptually independent of their publisher's background" (Falkenberg, 2010: 233) including experimental features rather than day-to-day accounts of news being published. This stood in contrast to being a 'newspaper online'.

Falkenberg describes 'newspapers online' as those newspapers which approached the web as a place to shuffle their best print content, with little variation. *The Daily Telegraph* took this approach when launching the 'Electronic Telegraph' in 1994, catering to a small audience of readers with a text-heavy website designed to avoid taxing time or bandwidth (Richmond,

2009). This also minimized risk, in case the web didn't take off, by only updating the site once a day, and rather than investing capital to create new web content, replicated its print content directly online. This allowed the *Telegraph* to stake a claim as the first British newspaper website, while maintaining the daily agenda of its print newspaper brand (Williams 2015). It also shows where some saw the early web as, more or less, a new iteration of print. Sonia Livingstone (drawing on Manuel Castells) describes these as "attempts of content-providers to re-impose hierarchical, print-based models of authoritative information" (2004: 10-11).

For a time, this approach was effective. Either by holding back content from the web to entice readers to buy the print version, or by limiting access to subscribers of paid-for offline editions, such news sites operated 'in service' to their offline counterparts and brands. In a similar fashion, when the *Economist* went online in 1994, its site was tailored for "a loyal following among people who were at least aware of *The Economist*" (*Economist*, 2012), but its website was primarily a portal to other interesting sites online and not content. Made on an $120 budget, bedecked with graphics scanned at a copy shop, only later did it provide online readership with *Economist* content. A 1996 archive notes its tentative move towards online publishing, calling it: "the on-line home for a limited selection of articles from the current issue" (economist.com 1996).

It is clear such approaches were also part of ongoing discussion among those moving online. Former New Media Lab director Ian Katz (2009) refers to the *Telegraph* as offering "efficient and well-used digital facsimiles of their print editions", contrasting "the intoxicating world of possibilities" the *Guardian* sought to engage. The New Media Lab shied away from this, though it wrestled as well with how much to publish freely on its sites, and what to keep in its print pages. This was not merely a debate of the late-1990s, however. In 2006 when *The Guardian* announced it would publish some news online first, it reflected again on this tension, opting to push breaking news content online as, "the paper aims to widen and deepen coverage online to benefit *The Guardian's* expanding global readership" (*Guardian*, 2006), without sacrificing the value proposition made to those paying to read the analogue newspaper:

> Some exclusive stories will continue to be held back for the newspaper to maintain the quality levels of the print version. The object is to not remain beholden to a 24-hour printing cycle and be beaten to important news by print and new media rivals. (Ibid.)

The website at guardian.co.uk at first bore a strong resemblance to 'online newspapers', and in both presentation and from accounts of those involved in developing these sites it was (in spirit, if not formally) conceptually independent of the paper it shared a name with. Only later did it more strongly reflect features of the newspaper itself, prioritising similar news agendas and reflecting its offline identity online. When the New Media Lab received a flood of complaints in 1997 for their lack of online coverage of Princess Diana's death they realized the extent to which the web was taking hold as a news medium. The New Media Lab

rebounded by offering online readers a special news section on Diana, advertising this online supplement under related print coverage (cf. Ahmed 1997), and treating the website from that point forward as a space where news content could also be sought. This was a familiar tail among news media, as more and more users were able to get online and news organizations sought to 'catch up' so their sites could keep apace of both news and web developments and in ways they hadn't previously prioritized (Allan 2006: 27). As this developed, guardian.co.uk – launched as 'Guardian Unlimited' – started to resemble the idea of 'newspapers online'. This second stage came in 1997, by Katz's account (2009), and solidified when Emily Bell took over digital strategies in 1999 (Thompson, n.d.).

These considerations point to where the *Guardian* is both an obvious and a complicated case to explore, and where its trials serve as key points for understanding journalism's contemporary history. First, while it has been lauded for adopting a digital-first strategy in the 21st century (Rusbridger 2010), its guardian.co.uk site was a relative latecomer among news media. Yet by launching years after other media had taken hold of the web's opportunities, it could adopt successes from the early web – including successes its own team, such as Bill Thompson, helped develop. This, Michael Stevenson argues, allowed news media going online to harness the developments of the early web, "years before it was accessible to mainstream audiences" (Stevenson 2016a: 1088), and to use the web in "ways that extend (rather than oppose) 'old media' values like branding and a distinctive editorial voice" (Stevenson 2016b: 1331).

Guardian.co.uk: A 'lateral webcentric spirit'
Marking the web cultures its developers came from, including its first New Media Lab director, Bill Thompson, in its early years the guardian.co.uk site featured unique, sometimes peculiar, content alongside news. This was part of a "lateral, webcentric spirit" embraced in the lab, where experimentation was prioritized over dogged commercialism (Katz 2009). As the product of "a new media 'skunkworks' team, tasked with dreaming up innovative online ideas", the New Media Lab tinkered with design elements constantly, and features on the page emerged and disappeared regularly, "like a child experimenting with every ringtone on a new phone" (Katz, 2009). Indeed, though the archived landing pages at guardian.co.uk show a relatively consistent template from day to day, its other pages make experimentation a feature. The webzine Shift-Control offers one such example, discussed below, with a new design for each edition that was released (*Wired*, 1997); its education pages hosted synopses of major world news for use in schools, as well as experimental discussion spaces (guardian.co.uk 1997a). News content was not presented in incremental updates, but took approaches such as providing a year in review with a navigable 'digital library' of news content from 1996 (guardian.co.uk 1997c). It also promoted commentary, including a broadside against the *Sun's* 'homophobic mafia' (guardian.co.uk 1998), and a campaign to save the BBC World Service (guardian.co.uk 1997b).

Its 'webcentricity' was one part of the picture, but archives also show where the demands of a profit-driven newspaper industry came into conflict with the non-commercial idealism of the web. Former New Media Lab director Bill Thompson recalls this:

> we were convinced we could change the world of newspapers, We [sic] were playing around with design, navigation, content and approach, looking for ways to take the printed newspaper online while preserving the values and attitude that made it work: we were all about brand extension, not about being an income-generating business unit. (Thompson n.d)

Such clashes between web and commercial priorities are to be expected at moments of convergence (Bruns 2009; Jenkins 2006), and the *Guardian* was not immune to these. When it comes to extending the brand, the guardian.co.uk page regularly promoted *The Guardian*'s offline content, including through a drop-down menu for sports and election news (under pages for 'Ashes 97' and 'Election 97'), but also its non-news, commercial projects. This includes *The Guardian's* job listings portal, RecruitNet, and its magazine *Auto Trade*r, which at one time was a prominent revenue stream for the *Guardian Media Group*. Both are prominently linked to on the main pages. The site also carried overviews – but not full accounts – of daily news coverage, with encouragement to find more by buying the printed paper (guardian.co.uk 1997d).

In the reverse, bringing readers from print to the web, *The Guardian* offered subscribers who had home delivery a bonus e-mail edition of *Guardian Weekly* publication, which pulled content from a global network of partners (cf. Guardian Weekly 1997a, 1997b). Elsewhere, the paper regularly urged readers to go online, running in-house ads for the landing pages at guardian.co.uk and go2.guardian.co.uk, and including the URLs at the end of stories where additional content was online. Its 'OnLine' print supplement, for instance, nudged readers to "COME and visit us at Go2, our Web site, where you can find links relating to articles in the latest issue of OnLine" (*Guardian* 1996).

The balance between print priorities and web innovation is also apparent in *Guardian* coverage of the internet in its pages. While in part it acknowledges the novelty of the web, we also see signs of now-familiar processes of "normalization" (Singer 2005), referring to the ways innovation developed online, outside traditional media, is later adopted by mainstream news to make 'normal' these new media approaches. One example of this begins in 1994 when Bill Thompson designed the 'FringeWeb' site for the Edinburgh Fringe, featuring "an early blogging/journal tool" and publishing reviews of the festival from *The Guardian* and *Observer*, "live each morning, often before people got the newspaper delivered" (Thompson n.d.). In the pages of *The Guardian*, FringeWeb is then mentioned in 1995, when Thompson and Dave Potts announce: "The *Guardian*, a sponsor of FringeWeb '94, is setting up its own site to complement the Fringe Society" (*Guardian* 1995). This site promises reviews, again, as well as links to the festival's programmes and its own official site. This could hardly be seen as lifting Thompson's idea – he wrote for *The Guardian* and helped design its first websites.

Instead, it reflects how newspapers were keen to identify innovative approaches developed by those at the vanguard of the web (like Thompson) to capitalise on these successes and accelerate the appeal of their own platforms.

As *The Guardian* was exploring the best ways to deal with the web as a counterpart to the print media object, it did so by embracing both its innovation and its novelty. In these years, *The Guardian* regularly ran a section titled 'Internet', publishing site listings sent in by its readers. These included brief descriptions of interesting sites – e.g., 'Planet of the Apes Timeline', featuring calendars for the timelines of the film series; 'Sold Out!', with a collection of Japanese adverts starring Western celebrities; or, 'Super marketing; Ads from the comic books', which publishes old mail-in adverts from comic books (*Guardian* 1997). 'Internet' also published daily diaries, much like cinema listings or an events calendar, which showed when users could go online for everything from discussions with authors to live acid jazz concerts, to "The Hub, Spiritual guidance" which readers could seek online, at 4.30 a.m. on 11 December, 1997 (*Guardian* 1997). Looking back, this seems disjointed. Why use a printed page to detail cumbersome URLs and web opportunities? However, at a time when audiences were moving between the printed page and digital screen unsure of which would hold sway in the end, it reflects elements of the discovery and experimentation of traditional media addressing the new medium of the web.

Part of newspaper strategies, including *The Guardian*'s, in experimenting with the web as a viable avenue for news meant bringing readers along to the web, showing those who may still be unfamiliar with the web its possibilities. In its pages, the *Guardian* regularly walked readers through its online experiments, and introduced new digital services it offered. For the former, a column titled "Live wire politics" by Vincent Hanna in 1996 encouraged readers to submit questions to an online debate he would host with Paddy Ashdown: "the first of an experimental group of online 'conversations' with the great and the good" (Hanna 1996). Hanna went on to detail the technology and process that would be followed to conduct this and future conversations (cf. Berlins 1996). Rather than treat the web as wholly distinct from the paper's efforts, this shows instead where *The Guardian* sought to build linkages to the web, and stay active in both print and digital domains. Yet the web was also a space for developing new services to offer readers, and experiment with new revenue opportunities. Several news articles in 1997 reflect this in their coverage of new types of job search web portals emerging online. Each of these articles also describes *The Guardian*'s own RecruitNet portal (Baker 1997; guardian.co.uk 1997d; Wylie 1997), which was one of several similar services newspapers established as revenue from printed job listings dissipated in the late 1990s (Picard, 2008).

Missteps in webcentricity
Processes of innovating on the web in these early years also meant missteps, regardless the path taken. Using the web as a cypher for print tired after a few years for the *Telegraph*.

Holding back content for paying readers has had, at best, mixed results for the British press (Arrese, 2016). Likewise, approaches taken by *The Guardian* had their own pitfalls which show it was not inured to the culture clashes that regularly erupted between "webcentric" employees (Katz, 2009) and those who came to the web from traditional newsroom. Nor was it insulated from the different priorities of the web for news and news as a business. Both Bill Thompson, who led *The Guardian's* New Media Team and its first sites (Thompson, 2009), and Ian Katz (2009) who later led the NML, describe *The Guardian*'s web projects as uniquely webcentric – and as "disruptive, open, and participatory". At times, however, this was more a romantic narrative than an observable reality (Stevenson, 2016b: 1088-1089). While Katz (2009) waxed nostalgic over the "curious range of experimental projects" the New Media Lab dreamt up, Thompson's accounts highlight tensions between the print paper and web sites. He separately denigrates Katz's leadership, and notes competing priorities in bringing *The Guardian* to the web: "I left *The Guardian* for many reasons, not least being the tension between the newsroom and its journalists and the Web team under my direction. Since leaving I can't understand why I stayed as long as I did" (Thompson n.d.). These are emblematic of the push and pull between control and creativity that defined the digital turn for many newspapers (Lewis 2012).

For *The Guardian*, the line between web innovation and web mistake was regularly transgressed. Short-lived experiments include an ambitious attempt at 'singing scores' for football results, though its Euro '96 football coverage set the standard for football news online. Among its more discussed missteps is a webzine, Shift-Control, that featured (among other things) the memoirs of a 40-year-old bee from Manchester named Freebee (Buzzin 1997). While this might seem woefully out of place for a newspaper site, it offers a useful illustration the socio-cultural dynamics that developed with the emergence of the web. It also shows where newspapers' approaches to the web sometimes ran counter to the nature of web and cyber cultures already established online.

Shift-Control published a diverse range of content weekly. Through plentiful hyperlinks, animated graphics, and interactivity, it sought to be a digital approximation of free-form analogue zines (Duncombe 1997). Shift-Control foregrounded its identity around cultural (and counter-cultural) content, rather than the *Guardian*'s branding. Moreover, it often intentionally elided the news. Its April 1997 was titled 'NOT THE ELECTION ISSUE', proclaiming: "This week Shift Control comes to terms with the electoral circus – by willing it out of existence" (Shift-Control 1997a). Shift-Control was innovative, though. Its readers could write their own cultural reviews on its 'Remote Control section' (Ibid.), or send an animated email from Buzzin' himself (1997b). Contests enticed readers to submit their own short stories (Shift-Control 1997c), and long before BuzzFeed popularized the feature, it featured an online quiz to help determine "How wild are you … about sleaze?", with tongue-in-cheek commentary and a chance to choose your favourite euphemism (Shift-Control 1997d). It bore all the markers of an experimental approach to publishing online, including

those more familiar to web cultures than to newspapers' occupational cultures (Stevenson 2016a).

From accounts of those behind Shift-Control, it wanted to carry the spirit of zines, those "non-commercial, non-professional, small-circulation magazines which their creators produce" (Duncombe 1997: 6), to a new online space. What was done with cheap mimeograph machines and copy shops for print, Shift-Control did digitally: "All the pages were hand-built, so you were not only pushing copy through", said editor Rada Petrovic (quoted in *Wired*, 1997). Not only was it recreated each issue, the way it was referred to as Shift Control, or Shift-Control, each riffing on keyboard commands, varied from reference to reference. Thompson, also interviewed by *Wired* described it as a project to show those in charge at *The Guardian* that "an online magazine could be really successful and make a splash" (Ibid.). Shift-Control did so by using *Guardian* resources, including the New Media Lab and web site, without resonating with the printed paper's journalistic ambitions. As an example of 'culture jamming' – subverting mainstream media by using traditional formats to offer alternative content (Dery 1993) – it offers a unique artefact of the way innovators seized on a digital moment, and newspapers grappling with how to make use of the web.

However, distinct from its predecessor 'zines, Shift-Control was *not* non-commercial. It was funded by £200,000 from Whitbread (brewery of Boddington's, among other beers) and advertisements for Whitbread beverages are prominent. It was also *not* non-professional, at least inasmuch as it was a *Guardian* project. In the end, it was mired in such contradictions. "The very close relationship between advertising and editorial led some to, unfairly, criticise Shift Control for being created with an advertiser rather than an audience in mind," reported *Campaign* magazine (MacMillan 1997). *Wired* (1997) noted it was "colourful and provocative" though "not successful in marshaling [sic] a large audience". Another perspective might say its innovation was in conflict with its provocations, particularly considering the ethos of the very audience it might hope to secure. Its requirement for users to register with email addresses "drew flames on Usenet" (Ibid.), the communication network popular among web-savvy users, and contravened the expectations of web and cyber cultures which grated against such data gathering.

That *The Guardian* failed to ever make use of what registration data it did secure could have mollified these particular complaints, but then *that* contravenes the overall purpose of a profit-oriented newspaper trying to secure new audience bases online which they could then pitch to advertisers. In the end, Shift-Control's demise exposed *The Guardian*'s "lack of immersion in the tightly woven online communities that could have spread the word about it" (*Wired* 1997), and perhaps the misfit nature of a countercultural webzine for a news-based organization. It lasted 50 issues before being shuttered in 1997.

The end of this experiment coincided with *The Guardian* turning attention towards projects it hoped would attract more reliably steady readers, moving guardian.co.uk from an online platform for experimentation (though it remained, in part, devoted to such things) towards approximating the *Guardian* newspaper online as 'Guardian Unlimited'. In 1997, then-editor Alan Rusbridger instructed the New Media Lab to develop a site that was more consistent with the paper's ambitions. "Instead of attempting to transfer the paper online, we would build deep specialist sites" Katz wrote, adding the goal was to engage "the new medium's strengths rather than simply recycling its print product" (Katz 2009). This was not embraced across the team – in response to the "instructions to redirect their efforts to building an online version of *The Guardian* itself, the team wore the despondent look of a bunch of German soldiers who had just been sent to the Eastern Front" (Ibid.). Despondent as they may have been, when we look overall at how *The Guardian* navigated the digital turn, the examples here show where among a range of web experiments and projects, *The Guardian*'s endeavours reflect the sort of media culture shifts that come naturally with moments of technological adaptation (Jenkins, 2006).

From there it took off, registering one million readers in 1999. By the start of the 21st century, it had become a prominent news website, winning an *Editor & Publisher* award for best design for a newspaper's 'online service' in 2000 at a time when those awards were still dominated by U.S. news websites (*Editor & Publisher* 2000). Its early stages from print to web outlined here show how *The Guardian* moved, "from experimental sandbox to the heart of what we do" (Katz 2009), where it now registers among the most sought English news sites online (Alexa.com 2017). In the context of newspapers embracing the web, experimentation was certainly key for *The Guardian* early though it was not alone among media in seeing the web as a space for trial balloons. That some of these, from blogs, to interactive fora, to deep specialist sites, have now become normalized in ways which make their approaches seem common now shows where the successes of newspapers going online have also shaped our expectations for newspaper content online.

Archiving the Digital Turn
As a discussion of challenge and opportunity, newspapers going online seem to tell a specific story of engaging with new platforms and media spaces. Their trials have included short-lived experiments, and more successful patterns of engagement. For newspapers' digital histories, the lessons from these transitions extend further. This final section will point to the challenges and opportunities for journalism historians, particularly as print archives have been digitized and as online news has become prominent. While in practice the priorities of forward looking news sites and history-focused archives are discrete, they overlap in ways that speak to the changing nature of information in a digital age. In instances these also show where the early web and its open approaches to content have introduced new debates for archives expanding their digital offering.

At the same time newspapers were trying to make sense of digital opportunities, libraries started to confront what these technologies might mean for their collections as the 20th century turned to the 21st (King 2005). The opportunities were clear, albeit challenging. Digitization meant converting analogue content to data files in the first instance, but also offered a chance to make vast print and paper collections more accessible to scholars (Steel 2014), thereby reinvigorating research libraries' missions. For the British Library (BL), digitization of newspapers began in 1992 with its most used microfilm archives, the Burney Collection, containing papers dating back to 1602. By 1995, 6,000 images per month, scanned from microfilm, were being stored (King 2005: 177). This was respectable, but insufficient when considering the scale of material to be processed, stored on 45 kilometres of shelves (Ibid. 166), so curators turned towards collaboration with private companies which were better resourced for the range and pace of digitization. In 2001, partnerships with software companies and research partners prioritized improving character recognition and indexing (and therefore searching), and later in 2004 the Joint Information Systems Committee (JISC) backed a project to scan 2 million 19th century newspaper pages, a collection now accessible through the BL's reading rooms and Gale Cengage's commercial platforms. Similar digitization efforts also took off around the UK; the national libraries of Wales, Ireland, and Scotland have similar projects underway, each with varying levels of access. The Wales collection launched in 2013 with an online accessible archive of its pre-1910 Welsh newspaper archives (National Library of Wales n.d., Tedd 2011)., while the National Library of Ireland launched its digitization project in 2010 (National Library of Ireland, n.d.), with access primarily limited to those at library or partner facilities.

Such partnerships helped build digital catalogues, including many which can be accessed through websites, even when most digitised material is only accessible within research libraries. While these greatly sped up the ability for public research libraries to create digital archives, they also posed challenges as archives and collaborations grew. Much like the uncertain approaches of website, with features emerging and disappearing from iteration to iteration, this mix of collaborators and libraries also meant a mix of how data was presented and accessed from project to project. For scholars embracing digital humanities tools such as keyword and metadata searches, this posed specific complications (Wijfjes 2017: 14). These projects also tended to focus on two specific goals: Securing print content in particular, mostly from before 1900, and improving the accessibility of these resources for researchers by cataloguing what was held. Thus, they tend to prioritize old content and developing robust catalogues. In part this is due to resources, as well as legal allowance. It also speaks to a wariness among newspapers who attribute the commercial downturn print newspapers saw in the early 21st century to the web taking off as a space for mostly-free content (Picard, 2008). While charging for current content for British papers has been a struggle, charging for archives remained viable as a market emerged for users interested in genealogy, for instance, and with more recent content where proprietors retain copyright they are reticent to relinquish these avenues. While national libraries have a legal right to retain copies of newspapers, they do

not under UK law have a right to make these public beyond their walls, a limit which extends to what can be archived of web content (UKWA, n.d.).

Indeed, much of what shaped the limits on content in the dead-trees newspaper age continue to affect online archives. Still, for historians these resources opened access to otherwise difficult to consult resources, and resources are constantly improving, in part because contemporary newspaper content coming into national deposit libraries are easily stored digitally, compared to those "created in the letterpress era, which require conversion to digital formats using scanning equipment" (King 2005: 167).

What these developments fall short of confronting, however, is the context of changing conceptions of what a newspaper is, as digitization meets digital news. From lessons learned exploring newspapers undergoing the digital turn, we can see archives exploring similar tensions between opening up and maintaining control as they confront a new type of archive – the digital news archive, which has lagged in development even as the web has grown. For years, libraries like the BL were able to experiment with, but not archive, web content at the scale they have been archiving printed works. This only changed in the UK in 2013, when The Legal Deposit Libraries (Non-Print Works) Regulations 2013 extended what national libraries could archive to include the web. This allowed the BL, and other national libraries, to archive UK websites, including blogs and news sites (Ashton 2018; UKWA, n.d.).

However, access to these archives remains limited to reading rooms and the BL's own computers with their download restrictions in place. While other web content archives – including those of the Internet Archive – are accessible freely online, these too are limited in what they capture and, in contrast, the new non-print deposit regulation allows the BL to build a stable collection of British websites which it has oversight of. While this is a recent development, the BL and others are not starting from a standing start, having previously partnered with organizations including JISC and the Internet Archive to create some records of what was being published online and derived data sets, which can also be explored in its reading rooms (BL.uk, n.d.). This is a welcome step for news researchers, though not without considerations. As Brügger (2010a, 2010b) notes, where digitized print content often offers a fair facsimile of what was once on page, web archives operate differently. For one, they are resource heavy and call for negotiating specific technical challenges, including how often to store pages, how deep to build an archive, and which specific elements to archive. As one example, content on the web does not follow a linear presentation of information, and from everyday web use it is clear that storing any archive of a site will need to confront hyperlinks that extend to subdomains, incorporating where possible interactive features, and increasingly wrestling with graphics and features which require more and more processing power. Nevertheless, for researchers the digital turn has meant a huge opening up of opportunities as the transitions of a digital age have meant greater access to

news resources, which now includes shifting from 'newspaper' curation to 'news' curation as web news archives are added to their catalogues (McKernan 2017).

Conclusion: The move from experimentation

The story that has emerged of *The Guardian* going online highlights the broader context of digital change for British newspapers and periodicals at the end of the 20th century, and in general shows where, for newspapers taking the digital turn, this was a period of trial and experimentation. In the overview of archives digitizing news, a similar case emerges that shows where – for British newspaper histories and historians – the digital turn has been one of specific complications that come with the increased access to digital technologies. For both newspapers and archives, the enthusiasm that came with the emerging World Wide Web from the 1990s onward colour the ways we consider digital and digitized news help to explain the specific dynamics of a disruptive moment within British media (Eldridge, 2015). The *Guardian*'s first steps online show a willingness to grapple with this new space in interesting ways, and later where these were fraught with challenges. The discussion here shows as well how this had a knock-on effect for archives, who face pressures both in making sense of digital technologies for their offline content, and for how they have adapted to new ways of accessing content. Considering the two, we can better understand how newspapers and their archives made sense of the digital turn and entering a digital age.

References

Ahmed, K. (1997) 'Editors bar snatched pictures of princes', *The Guardian,* 9 September*.*

Alexa.com (2017) 'The Top 500 sites on the web: By Category', accessed 30 September at: https://www.alexa.com/topsites/category/News

Allan, S. (2006) *Online News,* Maidenhead: Open University Press.

Arrese, A. (2016) 'From Gratis to Paywalls', *Journalism Studies*, 17(8): 1051-1067.

Ashton, L. (2018) 'Archiving the UK Copyright Literacy Blog', at: http://britishlibrary.typepad.co.uk/webarchive/

Baker, P. (1997) 'Log on to a world of job opportunities', *The Observer*, 24 August.

Berlins, M. (1996) 'Writ large', *The Guardian*, 10 December.

BL.uk (n.d.) 'Collections: UK Web Archive', at: https://www.bl.uk/collection-guides/uk-web-archive#

Brügger, N. (2010a) 'Web History, an Emerging Field of Study', in: N. Brügger (ed.) *Web History*, London: Peter Lang (pp. 1-25).

Brügger, N. (2010b) 'Web History: An analytical grid', in: N. Brügger (ed.) *Web History*, London: Peter Lang (pp. 29-59).

Bruns, A. (2009) 'From Prosumer to Produser: Understanding User-Led Content Creation', *Transforming Audiences 2009*, 3-4 Sep, 2009, London.

Buzzin (1997) 'Chapter one', *Shift-Control*, 31 July [archived at: https://web.archive.org/web/19970731203442/http://shiftcontrol.com:80/freebee/ch1free1.html]

Curran, J. (2012) 'Rethinking internet history', In: J. Curran, N. Fenton, and D. Freedman (eds.) *Misunderstanding the Internet*. Abingdon: Routledge (pp. 34–65).

Dery, M. (1993) 'Culture Jamming', *Open Magazine Pamphlet Series, #25.*

Duncombe, S. (1997) *Notes from Underground: Zines and the Politics of Alternative Culture*, London: Verso.

Economist (2012) 'Difference Engine: Lost in cyberspace', *Babbage Blog,* 9 July. At: https://www.economist.com/blogs/babbage/2012/07/digital-archives-0

economist.com (1996) 'Main Page', [archived at: https://web.archive.org/web/19961227080237/http://www.economist.com:80/]

Editor & Publisher (2000) 'EPpy Awards', at: http://www.eppyawards.com/Content/Past_2000_Winners-18-.aspx

Eldridge, S. (2015) 'Change and Continuity: Online media through history's lens', in: M. Conboy and J. Steel (eds.) *The Routledge Companion to British Media History*. Abingdon: Routledge (pp. 528-538).

Falkenberg, V. (2010) '(R)evolution under Construction: The Dual History of Online Newspapers and Newspapers Online', in N. Brügger (ed.) *Web History*, London: Peter Lang (pp. 233-254).

Foster, A. and Ford, N. (2003) 'Serendipity and information seeking: An empirical study', *Journal of Documentation*, 59(3): 321-340.

Guardian (1995) 'fringe online', *The Guardian*, 5 August.

Guardian (1996) 'OnLine [section], *The Guardian*, 4 July.

Guardian (1997) 'Internet', *The Guardian*, 6 December.

Guardian (2006) 'Guardian to offer news online first', 7 June, at: https://www.theguardian.com/media/2006/jun/07/theguardian.pressandpublishing)

Guardian (2010) 'History of the Guardian website', 1 November, at: https://www.theguardian.com/gnm-archive/guardian-website-timeline

guardian.co.uk (1995) 'Main Page', 5 November [archived at: https://web.archive.org/web/19961105173344/http://www.guardian.co.uk:80/]

guardian.co.uk (1996) 'Main Page', 22 December [archived at: https://web.archive.org/web/19961222060658/http://www.guardian.co.uk:80/]

guardian.co.uk (1997a) 'Education', 22 February [archived at: https://web.archive.org/web/19970222165311/http://www.guardian.co.uk:80/education.html]

guardian.co.uk (1997b) 'Save Bush', 22 February [archived at: https://web.archive.org/web/19970222170349/http://www.guardian.co.uk:80/savebush/]

guardian.co.uk (1997c) 'G96', 19 June, [archived at: https://web.archive.org/web/19970619001148/http://www.guardian.co.uk:80/g96/]

guardian.co.uk (1997d) 'Main Page', 19 June, [archived at https://web.archive.org/web/19970619000729/http://www.guardian.co.uk:80/]

guardian.co.uk (1998) 'Main Page', 11 November, [archived at:
https://web.archive.org/web/19981111185636/http://www.guardian.co.uk:80/]

Guardian weblog (2000) 'Guardian weblog', 17-28 April, [archived at
https://www.theguardian.com/weblogindex/0,,340346,00.html]

Guardian Weekly (1996) 'gweekly', 22 February [archived at:
https://web.archive.org/web/19970222165319/http://www.guardian.co.uk:80/gweekly/]

Guardian Weekly (1997) 'gweekly', 19 June [archived at:
https://web.archive.org/web/19970619001204/http://www.guardian.co.uk:80/gweekly/]

Hanna, V. (1996) 'Live wire politics', *Guardian*, 7 March.

Katz, I. (2009) 'The Guardian's First Forays on the Internet', *The Guardian*, 23 October.

King, E. (2005) 'Digitisation of Newspapers at the British Library', *The Serials Librarian*, 49(1-2): 165-181.

Jenkins, H. (2006) *Convergence Culture,* New York: NYU Press.

Lewis, S.C. (2012) 'The Tension Between Professional Control and Open Participation', *Information, Communication & Society,* 15(6): 836–866.

Livingstone, S. (2004) 'Media literacy and the challenge of new information and communication technologies', *Communication Review*, 1(7):3–14.

MacMillian, G. (1997) 'Whitbread to carry on with Guardian', *Campaign,* 30 November, at:
http://www.campaignlive.co.uk/article/whitbread-carry-guardian/124482

McKernan, L. (2017) 'Analysing the past', at:
http://blogs.bl.uk/thenewsroom/2017/01/analysing-the-past.html

McKie, R. (1995) 'Britain Takes to Fast Lane of Superhighway', *The Guardian,* 26 March.

National Library of Ireland (n.d.) 'Digitisation', at https://www.nli.ie/en/digitisation.aspx

National Library of Wales (n.d.) 'Welsh Newspapers Online', at:
https://www.llgc.org.uk/index.php?id=4723

Picard, R. (2008) Shifts in Newspaper Advertising Expenditures and their Implications for the Future of Newspapers', *Journalism Studies*, 9(5): 704-716.

Richmond, S. (2009) 'Telegraph.co.uk: 15 years of online news', the *Daily Telegraph*, 11 November, at: http://www.telegraph.co.uk/technology/6545788/Telegraph.co.uk-15-years-of-online-news.html

Rusbridger, A. (2010) 'The Hugh Cudlipp lecture: Does journalism exist?', *The Guardian,* 25 January, at: https://www.theguardian.com/media/2010/jan/25/cudlipp-lecture-alan-rusbridger

Shift-Control (1997a) 'Issue 26', *Shift-Control*, 11-17 April [archived at:
https://web.archive.org/web/19970416092235/http://shiftcontrol.com:80/]

Shift-Control (1997b) 'It's a Freebee', *Shift-Control*, 16 April [archived at:
https://web.archive.org/web/19970416092526/http://shiftcontrol.com:80/freebee/freebee.card]

Shift-Control (1997c) 'White Lies', *Shift-Control*, 16 April [archived at: https://web.archive.org/web/19970416092749/http://shiftcontrol.com:80/white-beer/whitelies2.html]

Shift-Control (1997d) 'Sleaze Quiz', *Shift-Control*, 16 April [archived at: https://web.archive.org/web/19970416092857/http://shiftcontrol.com:80/quiz26/quiz.html]

Singer, J.B. (2005) 'The political j-blogger: 'normalizing' a new media form to fit old norms and practices', *Journalism,* 6(2): 173–198.

Stevenson, M. (2016a) 'The cybercultural moment and the new media field', *New Media & Society,* 18(7): 1088-1102.

Stevenson, M. (2016b) 'Rethinking the participatory web', *New Media & Society,* 18(7): 1331-1346.

Tedd, L. (2011) 'People's Collection Wales: Online access to the heritage of Wales from museums, archives and libraries', *Program*, 45(3): 333-345

Thompson, B. (n.d.) 'new media dog', *the billblog* at: http://www.thebillblog.com/about-bill/new-media-dog/

Thompson, B. (2009) 'Those who do not understand the past…', *the billblog,* 30 October, at: http://www.thebillblog.com/2009/10/those-who-do-not-understand-the-past/

Toms, E.G. (2000) 'Understanding and facilitating the browsing of electronic text', *International Journal of Human Computer Studies*, 52: 423-452.

UKWA (n.d.) 'About', at: http://www.webarchive.org.uk/ukwa/info/about

Wijfjes, H. (2017) 'Digital Humanities and Media History: A Challenge for Historical Newspaper Research', *Tijdschrift voor Mediageschiedenis*, 20(1): 4-24.

Williams, R. (2015) 'The early 25 days of websites', *Daily Telegraph,* 21 December, at: http://www.telegraph.co.uk/technology/internet/10663451/The-early-days-of-25-websites.html

Wired (1997) 'UK's *Guardian* Newspaper Closes Webzine', *Wired*, 30 September, at: https://www.wired.com/1997/09/uks-guardian-newspaper-closes-webzine/

Wylie, I. (1997) 'Net is biggest and best for job adverts', *The Guardian*, 6 December.