

## **Digitising the past: Next steps for public-sector digitisation**

Alastair Dunning

Digitisation Programme Manager, JISC (Joint Information Systems Committee)<sup>1</sup>

London, United Kingdom

July 2009

Barely a week goes by without news of the digital publication of some sparkling cultural resource being splashed across the media - 2009 has been a particularly rich year.

"World's oldest Bible published in full online", reported the Daily Telegraph in July 2009, as the website with the Codex Sinaiticus was introduced to the general public.<sup>2</sup> "Frozen in time: Historic images of polar exploration made public", proclaimed the Daily Mirror, as the University of Cambridge published its stunning collection of images of early voyages to the Arctic and Antarctic.<sup>3</sup> And the Independent marked the launch of the University of Kent's online archive of twentieth-century political cartoons by reporting on "a century of satire, wit and irreverence".<sup>4</sup>

Beyond the early blaze of publicity, some digitisation projects have proved wildly spectacular, maintaining a considerable audience over time. The Complete Work of Charles Darwin Online, which provides access to numerous digitised documents relating to the naturalist (including several editions of *The Origin of the Species*), claimed 90 million hits between October 2006 and July 2009.<sup>5</sup> The Old Bailey Online, giving extensive details of criminal trials in London's central criminal court from 1674 to 1913, averages 6,000 unique visitors a day, which means yearly visitors of just under 2.2m.<sup>6</sup>

More notoriously, the success of some sites in capturing widespread interest has caused some teething problems. The Vision of Britain site, which provides historic and social information on every place in Britain, temporarily crashed after its creator appeared on breakfast TV to promote the site.<sup>7</sup> And the same fate befell the National Archive's 1901 Census site, a fact that really awoke the organisation to the huge international interest to the rich seams of genealogical information The National Archives holds, and proved a springboard for some spectacular digital successes.<sup>8</sup> This was only a temporary blip for the Vision of Britain site too - it now regularly achieves 80,000 unique users a month.<sup>9</sup>

A casual observer, therefore, might think that this digital world is a well-constructed and ordered world. Certainly, the successes mentioned indicate so and deserve praise. Yet the early optimism in digitisation has taken something of a battering, and revealed a slightly more anarchic digital sphere. The birth of the Internet spawned a tiresome flood of clichés and suppositions about presenting digital information online - the belief that knowledge could be instantly democratized via global internet access, that digitisation projects were reasonably simple DIY jobs requiring only basic training, that publishing would be cost-free

---

<sup>1</sup> The JISC is the funding body for supporting the use of ICT (Information and Computing Technologies) within the UK Higher and Further Education Sectors

<sup>2</sup> <http://www.telegraph.co.uk/news/newstopics/religion/5749555/Worlds-oldest-Bible-published-in-full-online.html>. The manuscript is available from <http://www.codexsinaiticus.org/en/>

<sup>3</sup> <http://www.mirror.co.uk/news/top-stories/2009/03/04/frozen-in-time-historic-images-of-polar-exploration-made-public-115875-21170538/>. The polar photographs are available at <http://www.freezeeframe.ac.uk>

<sup>4</sup> <http://www.independent.co.uk/news/media/online/a-century-of-satire-wit-and-irreverence-986844.html> The Cartoon Archive is available at <http://www.cartoons.ac.uk/>

<sup>5</sup> Figures cited at the home page of <http://darwin-online.org.uk/>

<sup>6</sup> Tim Hitchcock, Robert Shoemaker, Digitising History from Below, <https://uhra.herts.ac.uk/dspace/bitstream/2299/38/1/102846.pdf>. The Old Bailey site is available at <http://www.oldbaileyonline.org/>

<sup>7</sup> The site is available at <http://www.visionofbritain.org.uk/>

<sup>8</sup> <http://www.1901censusonline.com/> Details about the crash are at <http://news.bbc.co.uk/1/hi/uk/1749045.stm>

<sup>9</sup> Unpublished project report for Historic Boundaries of Britain project, page 1.

via digital means, and that audiences, safely ensconced in their own homes, would flock to see these glittering jewels online.<sup>10</sup>

In the context of publicly funded digitisation projects, bitter realities soon disproved many of these innocent assumptions. In particular, the issue of sustainability - the fact that complex websites require ongoing funding in order to ensure their technical reliability and their intellectual freshness - was the unpleasant jack-in-the-box that kept popping up. Numerous projects received fixed term funding to digitise, say, their collection of historic architectural photographs and construct a relevant web-enabled database, only to find they had no funds to tweak and maintain the database, thus threatening its online existence. The result was an unhealthy glut of error 404 pages.<sup>11</sup>

Much of the detail of these problems was highlighted in a 2007 report 'Digitisation in the UK', produced by the University of Loughborough. The report cites an "impressive accumulation of a body of digital material" but points specifically to "deep fragmentation in all components of the digitisation infrastructure."<sup>12</sup> It offered a range of recommendations, including a UK framework for digitisation, the co-ordination of existing services and much greater understanding of user interaction with digitised resources.

Despite the clear direction provided by the Loughborough report, disenchantment seeped into the thinking of the funding bodies that had previously supported digitisation. Perhaps the most notable example was the cessation of the Arts and Humanities Research Council's (AHRC) Resource Enhancement scheme. Running from 2000, the scheme had supported 186 projects to make accessible resources and scholarly information in digital form, including both the Darwin and Codex Sinaiticus projects mentioned earlier.<sup>13</sup> But the scheme was abruptly stopped in 2007, the review stating that it considered 'the scheme to be ineffective in identifying and addressing gaps in resource provision and meeting the resource needs of the arts and humanities research community'.<sup>14</sup>

Similarly, digitisation was pushed down the agenda of Heritage Lottery Fund (HLF) and the Museums, Libraries and Archives Council (MLA); this may well have been tempered by the experiences of the New Opportunities Fund Digitisation Programme (commonly called NOF-digi).<sup>15</sup> Again, there were notable successes here, but despite a glossy review that highlighted them, there appeared to have been the feeling that the programme was something of a wasted opportunity; digitisation was a complex problem that needed much sophisticated strategic thinking before there could be a serious injection of further public funds.<sup>16</sup>

---

<sup>10</sup> This report largely focuses on the UK but the experience has been much the same elsewhere. The Shifting Gears essay sums up the US experience – "A lot of money went toward creating barely visited web sites. And a lot of institutions created preservation-quality images that they, in fact, had no way to sustain in the long run." Ricky Erway and Jennifer Schaffner. *Shifting Gears: Gearing Up to Get Into the Flow*. Report produced by OCLC Programs and Research, 2007, <http://www.oclc.org/programs/publications/reports/2007-02.pdf>, page 8

<sup>11</sup> The 404 error is the message given by a browser when it can no longer find a webpage. [http://en.wikipedia.org/wiki/HTTP\\_404](http://en.wikipedia.org/wiki/HTTP_404)

<sup>12</sup> University of Loughborough, *Digitisation in the UK*, 2005, <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/reports/digiukframework.aspx>

<sup>13</sup> A spreadsheet of these projects is available <http://hdl.handle.net/10760/17517>

<sup>14</sup> <http://www.ahrc.ac.uk/FundedResearch/Pages/ResourceEnhancementSchemeReview.aspx>. The full report does not appear to be publicly available. Other relevant research in the area is available from [http://www.ahrcict.rdg.ac.uk/activities/strategy\\_projects/reports/index.htm](http://www.ahrcict.rdg.ac.uk/activities/strategy_projects/reports/index.htm)

<sup>15</sup> There's a useful blog post on this at <http://bridgetmckenzie.blogspot.com/2008/05/mla-and-hlf-views-on-21st-c-curation.html>

<sup>16</sup> For a summary of the programme see Susi Woodhouse, "The People's Network and the learning revolution: building the NOF digitise programme", *Ariadne*, 2001 (29), <http://www.ariadne.ac.uk/issue29/woodhouse/>. The final evaluation report, published by consultants Education for Change in 2006 is at [http://www.biglotteryfund.org.uk/er\\_eval\\_ict\\_final\\_rep.pdf](http://www.biglotteryfund.org.uk/er_eval_ict_final_rep.pdf)

At the time of writing, the JISC (Joint Information Systems Committee - a UK body that supports the use of ICT in higher and further education) is the only UK organisation that has provided systematic funding for digitisation within recent years. In particular Phase 2 of its Digitisation Programme, which ran from 2007-9, learnt from many of the mistakes of the past.<sup>17</sup> Whereas earlier programmes had given projects funding and then left them to get on with it, the JISC provided much more strategic direction and funding was conditional on teams addressing each aspect of the digital life cycle. Thus JISC insisted that each project have sustainability plans in place so as to ensure longevity in the created resources; that each project undertake user analysis and implement marketing and communication plans; and more generally that each project be part of a larger digitisation community that could share, advise and learn from each other. The result has been a compelling set of digital resources that seem to be providing a much more systematic return on the initial funding.<sup>18</sup> Time will tell if this approach actually delivers the benefits it aims for.

Whatever the long-term success or failure of the JISC programme it is clear that digitisation involving public-sector funds requires a far greater degree of high-level strategic implementation than has been present previously. An article in the Guardian, looking at digitisation in the broader context of all content on the Internet surmised it nicely with the headline "Why a Wild West approach just won't do".<sup>19</sup>

There are encouraging signs, however, that the complexity of the digitisation jigsaw is being considered and pieced together, both in the UK and abroad. The *Shifting Gears* report, produced by the American library service OCLC, alerted the library and information world to the need for changed thinking in the area, pushing forward a provocative series of ideas that challenged previous practice in the area.<sup>20</sup> In the southern hemisphere, interesting work in both Australia and New Zealand is taking these agendas forward. In particular, the Digital NZ initiative is providing a strategic focus for all kinds of digital content within New Zealand.<sup>21</sup> And EU funding has also been directed to the development of infrastructure, rather than the digitisation of content. For instance, the Impact project is creating a European network of expertise so as to provide a platform for navigating the issues relating to the digitisation of text.<sup>22</sup>

Within the UK there are also similarly moves afoot. On behalf of the MLA, the Collections Trust are developing CultureGrid, as a way of harmonise existing services and content within the library, museum and archive community.<sup>23</sup> More broadly, the Strategic Content Alliance, a network of UK public sector bodies who share these concerns about the long-term issues of access to digital content, is tackling the key issues, and encouraging the implementation of the necessary policies.<sup>24</sup>

---

<sup>17</sup> The list of 22 projects funded by JISC within Phase 1 (2004-7) and Phase 2 (2007-9) of its Digitisation Programme are at <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/projects.aspx>. A full list of all digitisation projects funded by JISC since 2004, including its 2008-9 Enriching Digital Resources strand (which provided short bursts of funding to enhance existing digital content) is at <http://hdl.handle.net/10760/17520>

<sup>18</sup> Despite these successes, the economic circumstances of 2009 have meant there has not been a third phase to the Programme. (<http://digitisation.jiscinvolve.org/2009/07/03/does-jisc-have-funds-for-more-digitisation/>) However, JISC is continuing its support in the area, providing funding for projects to enhance their existing content or develop the necessary infrastructure for institutions to undertake their own digitisation.

([http://www.jisc.ac.uk/fundingopportunities/funding\\_calls/2009/02/grant0209econtent.aspx](http://www.jisc.ac.uk/fundingopportunities/funding_calls/2009/02/grant0209econtent.aspx))

<sup>19</sup> <http://education.guardian.co.uk/librariesunleashed/story/0,,2274844,00.html>

<sup>20</sup> Elway, op.cit.

<sup>21</sup> <http://www.digitalnz.org/about>

<sup>22</sup> <http://www.impact-project.eu/home/>

<sup>23</sup> <http://openculture.collectionstrustblogs.org.uk/2009/04/27/culture-grid-starts-to-take-shape/>

<sup>24</sup> The SCA comprises a number of primary stakeholders (the British Library, the NHS, the BBC, Becta, JISC and the MLA) and a larger number of secondary stakeholders. <http://www.jisc.ac.uk/contentalliance>

Perhaps most valuably, a report and series of case studies commissioned by this alliance have begun to shine some light on the bedevilling issue of sustainability. As mentioned above, cultural heritage institutions have struggled with how the staff and the financial support can be found to maintain online resources once the initial project funding has ended. The case studies are particularly enlightening, giving examples of projects, institutions and consortia that have developed the necessary skills and resources to survive in the digital world. This includes the Centre for Computing Humanities at King's College London, where a shared technical and human infrastructure has allowed for the deployment of over 80 separate websites of cultural and scholarly content.<sup>25</sup> There are further case studies on, for example, The National Archives' engagement with commercial partners, the harmonisation of free and commercial access strategies at the French Institut national de l'audiovisuel, and Cornell University's eBird website, that has harnessed the enthusiasm of amateur ornithologists to create a sustainable resource that meets the needs of both scientists and enthusiasts.<sup>26</sup> In all, the suite of twelve case studies present examples of the more sophisticated approach that will be required to support and sustain digitised content in the future.

-----

The work of the Strategic Content Alliance and others has been a response greater strategic harmony in the infrastructure that supports public sector digitisation. Yet as a community interested in continuing to explore what digitisation can offer, we need to be careful not to instil too much order. While anarchy might be too strong a term, a good dose of breaking established rules and practices will be an integral part of keeping digital content invigorated, helping reach out to diverse audiences on a global level.

It's worth remembering that earlier digitisation programmes were not entirely without a sense of order. Applicants to the AHRC's Resource Enhancement scheme were obliged to fill out a technical appendix, which asked for specific information about the technical standards to be used. Those aiming to use proprietary formats (e.g. creating master files in PDF or Microsoft Word) or planning to digitise at an insufficient standard (e.g. photographs as low resolution JPGs) were asked to rewrite plans to adapt a more open standards approach. Similarly, projects involved in the NOF programme underwent accessibility tests to ensure that their data and websites were being created in accordance with a long list of guidelines developed as part of the programme.<sup>27</sup>

Such an approach undoubtedly had considerable benefits. Data was captured in platform-neutral formats, thus minimising the chance of images and text becoming trapped in dated software. Websites were presented in a way that did not block access to those who utilised browsers other than the familiar Internet Explorer.

But there were drawbacks as well. Such control, particularly in the NOF-digi programme, demanded excessive documentation. More importantly, it placed partial restrictions on

---

<sup>25</sup> [http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/case-studies/SCA\\_BMS\\_CaseStudy\\_CCH.pdf](http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability/case-studies/SCA_BMS_CaseStudy_CCH.pdf)

<sup>26</sup> All twelve case studies are available from <http://www.ithaka.org/ithaka-s-r/strategy/ithaka-case-studies-in-sustainability>

<sup>27</sup> The original guidelines were present on the People's Network site, but the URL

(<http://www.peoplesnetwork.gov.uk/content/technical.asp>) no longer functions. The support service run by the UKOLN and the Arts and Humanities Data Service had a (still functioning) website at <http://www.ukoln.ac.uk/nof/support/>

those wishing to undertake innovative work and exploit formats that did not fall into the 'canon of openness'.

But since then there has been much discussion as to the extent to which rigid boundaries within which digitisation projects should capture and deliver their data. As standards guru Brian Kelly has documented, an emphasis on standards is to be applauded, but an over emphasis on standards ignores the limitations of such a framework.<sup>28</sup> Open standards can be costly, not mature or, as perhaps happens most often, run against the grain of the services and functionality a general audience expects to see.

The *Shifting Gears* report raised this argument another level, focussing not just on standards but the whole 'best practice' approach to digital capture. Best practice in digitisation has traditionally demanded careful attention to high resolution capture of rich digital images or other documents, with detailed hand-crafted metadata lovingly added in as context. To sum up the bullish response of *Shifting Gears* to this: "Quality vs. quantity—Quantity wins!" The report attributed the marginal position of digitisation within library infrastructure, and the under-use and subsequent sustainability problem of digital resources as caused by the overly precise methods – "Our intricate attempts to describe and present a few choice collections have resulted in expensive, but little-used websites. And the rest of our [non-digitised] collections remain largely invisible."<sup>29</sup> To the authors of *Shifting Gears*, slavishly following prescribed guidelines has got in the way of the bigger picture.

But it is not just best practice but a whole conceptual approach to digitised content that is still acting as a significant barrier to its wider diffusion. The overwhelming majority of digitisation projects have focussed on digitising their photographs, drawings, postcards or documents with the intention of delivering them from a custom made website, which they have been responsible for either creating themselves or, at the very least, paying somebody else to create. The result of this has been to create digital silos – lumps of digital content that cannot be shared, re-used or cross-searched without considerable difficulty, even with content of a similar nature.

To counter this, there need to be greater openness, allowing data to be shared, re-used and republished often for purposes different from those for which they were intended. Beyond the sheer quantity of documents that they can digitise, the clear advantage of the Google Books Digitisation Programme is that it places the digital content in a web environment (that of Google Search) with which there are already millions of users who are familiar with how the interface operates.<sup>30</sup> There is much greater of a user finding rare books from, for instance, the University of Oxford if they are hosted on a Google website rather than hidden on the university's ox.ac.uk domain.

In many cases, it is quite simple things that restrict the flow of content. For example, the British Library's newspapers digitisation programme has created a fantastic resource, which, by the end of 2009, should be presenting nearly 4m pages to the higher and further educational community, and around 2m pages to the general public.<sup>31</sup> Yet not only are the URLs which define each newspaper page lengthy affairs difficult to copy and paste, they alter according to the university or college where the user is hailing from. Thus if a lecturer

---

<sup>28</sup> B.Kelly, M.Guy, A.Dunning, *Addressing The Limitations Of Open Standards*, 2007  
<http://www.archimuse.com/mw2007/papers/kelly/kelly.html>

<sup>29</sup> Elway, op.cit., page 8

<sup>30</sup> <http://books.google.com/>

<sup>31</sup> <http://www.bl.uk/britishnewspapers>



from University of A wishes to send a reference to a news page to a colleague at B College, he has no immediate way of doing so.

Perhaps more importantly, floating, unstable URLs disrupt the process by which search engines analyse and index individual web pages. In light of the fact that the majority of users will search for content not by typing a URL into the address bar, but by typing a couple of search terms into Google, blocking out access to search engines constitutes a wasted opportunity to assist the discovery of a resource.

With sufficient technical planning, such problems as those cited above can be avoided. A more fundamental problem involves the sharing and cross-searching of related digitised resources that exist in different places. For nearly any topic one cares to choose, one can find numerous digitised resources of relevance scattered across the Internet.

A good example is posters relating to the first and second world wars. Posters provide an obvious target for digitisation given their fragility, obvious visual appeal, and historical importance as primary sources. A significant number of cultural institutions have therefore initiated projects that have included the capture of posters in their collections. They include the Imperial War Museum, the University of Minnesota, the University of Oxford, McGill University, the University of Washington and many, many more.<sup>32</sup> Yet there is no convenient way to search across these all these collections of posters; indeed, for many users, they may not even know of the existence of these digital collections. And take any other area – medieval manuscripts, early photography, documentary films, architectural drawings – and painstaking trawling via Google will identify a sprawling archipelago of thematically related digital content.

This is not a new problem. There have been numerous efforts to solve the conceptual issues in bringing together such rich content, often around the development of portals which either harvest or directly import metadata and then tailor this metadata to provide a functioning cross-search facility.<sup>33</sup> Such portals have met with partial success, but they have also been expensive involving plenty of staff time in bringing content partners on board, and then the laborious process of squeezing the metadata into the shape required for cross-search functionality.

To begin to overcome this issue content providers need to become much more promiscuous in how they let users gain access to their content. As mentioned above, previous methods of delivery have focussed on delivering the content via the organisation's own website, and perhaps also via a portal in the vein mentioned above. But such mechanisms still restrict the contexts in which users understand and manipulate the content.

There now exists a suite of different tools and standards, of varying level of complexity, which allow for cultural content to be harvested, imported, reused and re-visualised by global audiences. Content providers should no longer try and confine themselves to one standard by which they are making their content available, but embrace a range of lightweight technologies.<sup>34</sup> Such technologies include RSS, JSON and microformats, and

---

<sup>32</sup> A good list is at <http://www.oucs.ox.ac.uk/ww1lit/education/online/propaganda.html>

<sup>33</sup> The Vision and Sound Portal, based at the EDINA data centre in Edinburgh, is bringing together visual and audio content for usage in a HE context. A useful final report highlights some of the key obstacles in creating such a portal is at <http://edina.ac.uk/projects/vsmportal/docs/VSMFinalReport.pdf>. In the United States, the Aquifer initiative has looked at tackling many of the same issues (<http://www.diglib.org/aquifer/>). There are many others.

<sup>34</sup> Mike Ellis' 2009 presentation "*Don't think websites, think data*" is an excellent recapitulation of the issue. <http://www.slideshare.net/dmje/dont-think-websites-think-data>

richer standards such as RDF, which form part of the network of ideas that mesh together as the semantic web. Perhaps most importantly, the rise of the API (application programming interface), allows other users to build services and tools based on one's content.<sup>35</sup> The website Flickr is an excellent example. Originally conceived as a website allowing for the upload of photographic content, Flickr's installation of an API allowed for a host of subsidiary services to exploit this API and build services around it. Users can now create postcards, posters, books and other personalised goods, based on the content they have uploaded.<sup>36</sup>

Cultural heritage institutions are now starting to explore the advantages that can be gained from APIs. The Brooklyn Museum in New York has been one organisation pushing at the boundaries of what has been possible, and its API has allowed for the creation of numerous services that exploit its digital content in different ways.<sup>37</sup> The BBC Backstage project has also focussed attention on how developers and others can make use of BBC digital data.<sup>38</sup> Within the academic sphere, the Concordia project, a shared piece of work between King's College London and New York University, serves as a good example. Under development at the time of the writing, the project is exploiting lightweight geographical standards, notably GeoRSS feeds and KML, to integrate different data resources related to the study of classical Africa, mainly inscriptions carved in monuments created during the Roman Empire.<sup>39</sup>

Stopping digital content from flowing out to other locations is just one of the traditional boundaries digitisation projects need to be break down. How users interact with the digital data is another. Nearly all digitisation projects have followed a model where the place of knowledge (whether it be university, museum, archive or other) creates the digitised resource, adds the necessary contextual information and then presents it to a specific audience for consumption – a model that mimics the one-way flow of information in the lecture hall.

Successful Web2.0 websites, most notably YouTube and Flickr, have built their foundations on ignoring this model, allowing user-generated content to become the keystone around which their service is delivered. If they are to achieve greater success, future digitisation ventures need to question their traditional parameters, and develop services that respond to the greater demand for user interaction.

Within the UK, perhaps the most successful example of this has been the Great War Archive, an initiative that was borne out of a larger project, based at the University of Oxford, to digitise poetry manuscripts related to the First World War.<sup>40</sup> While the poetry manuscripts were held at either via collecting institutions (such as the Imperial War Museum's Isaac Rosenberg collection) or in private estates (such as the manuscript of Wilfred Owen's poem *Strange Meeting*), the Great War Archive sought to digitise relevant items held by the population at large. Members of the public were either invited to submit items in digital form, or to attend special collection days at local libraries, where their

---

<sup>35</sup> [http://en.wikipedia.org/wiki/Application\\_programming\\_interface](http://en.wikipedia.org/wiki/Application_programming_interface)

<sup>36</sup> <http://www.flickr.com/do/more/>

<sup>37</sup> <http://www.brooklynmuseum.org/opencollection/api/> The news link from this URL gives examples of the different ways in which the API has been exploited, e.g. creating applications for the iPhone, or various different interfaces which look at particular aspects of the collection.

<sup>38</sup> <http://backstage.bbc.co.uk/>

<sup>39</sup> <http://concordia.atlantides.org/>. Information on the standards is available at <http://en.wikipedia.org/wiki/GeoRSS> and <http://en.wikipedia.org/wiki/KML>

<sup>40</sup> The First World War Poetry Digital Archive is available at <http://www.oucs.ox.ac.uk/ww1lit/>, and the Great War Archive is embedded within this at <http://www.oucs.ox.ac.uk/ww1lit/gwa/>. The Flickr group for the Great War Archive is at <http://www.flickr.com/groups/greatwararchive/>

relevant material was photographed. All the material was then uploaded to the website where it could be searched for in context, or separately from, the material related to the war poets. Additionally a Flickr group was constituted that allowed for further dissemination of the content gathered via the Great War Archive initiative.

The result was overwhelming – in a few months over 6,500 items were collected including, according to the project website, “diaries, photographs, official documents, and even audio interviews with veterans”, with hundreds more added to the Flickr group. As well as proffering their own collections to the website, the public were also asked to contribute to the metadata, often bringing unique knowledge related to an item not to be found elsewhere. But perhaps of even more value was the broad community of interest that developed around the content, using the websites, joining the Flickr group, commenting on items, and, in many cases, supplying authoritative expertise on the collections presented.

An interesting summary of the project also reveals another eye-catching fact about such digitisation by the general public. The report notes that what “this initiative made clear was the potential for economies of scale that tapping into the potential for mass amateur digitisation could produce.”<sup>41</sup> Whilst each item in the ‘official’ Poetry Archive cost £40 to develop, digital items from the Great War Archive cost £4 each to create. The report continues: “Whilst the quality of the items in the Great War Archive could often be questionable, these figures do support the notion that further investigation is warranted to assess the possibilities of engaging the public directly to build community collections that are of educational and historic value.”<sup>42</sup>

Another notable example of user engagement comes from the National Library of Australia’s *Historic Australian Newspapers, 1803 to 1954*.<sup>43</sup> The project is tackling the problem of the poorish results of the automated conversion of the newspaper text from physical to digital form (commonly known as Optical Character Recognition, OCR) by inviting members of the public to correct the machine generated transcriptions. One may be sceptical of the enthusiasm by which ‘an ordinary user’ would relish such a task but there has been an immediate and rapid enthusiasm from local historians, keen to study and amend newspapers relating to their own area of interest, often based on particular people or places. The website’s ‘Text Correctors Hall of Fame’ not only gives evidence as to the popularity of the task but helps create a gentle sense of rivalry between transcribers.<sup>44</sup>

The dramatic success of these projects should persuade those digitising resources to the advantages that can be leveraged with such an approach. It allows for the collation of more material, and engages familiar and new audiences in novel ways. More directly, it should point them to the fact that there is obviously general enthusiasm to engage with culture on a digital sphere. This, of course, does not mean that the digital should replace the physical institution. Neither should such work be undertaken without due consideration of how it affects the institution’s image as a trusted place of learning, nor exaggerate the extent to which there is educational demand for information relating to individual family’s

---

<sup>41</sup> Kate Lindsay, *First World War Poetry Archive Final Report*, 2009

<http://www.jisc.ac.uk/media/documents/programmes/digitisation/ww1finalreport.pdf>, page 21

<sup>42</sup> Ibid.

<sup>43</sup> <http://ndpbeta.nla.gov.au/>

<sup>44</sup> <http://ndpbeta.nla.gov.au/ndp/del/hallOfFame>. The Old Bailey Online project has also developed a Wiki ([http://www.hrionline.ac.uk/obp-wiki/index.php/Main\\_Page](http://www.hrionline.ac.uk/obp-wiki/index.php/Main_Page)), which allows users to add contextual information, for example “biographical material about individuals and families who are documented in the Old Bailey Proceedings and Ordinary’s Accounts”.



histories.<sup>45</sup> But such methods of user engagement indicate the innovative ways in which digitised content can enhance the cultural and learning experience that universities, libraries and the like area are trying to provide.

It is perhaps a little too strong to tag these concerns for opening up data, breaking down boundaries and locating new users to be a call for anarchy. At best, it is call to dissolve some of the frontiers that digitising institutions have inherited whilst thinking about their collections as physical entities. Rather than keep it locked in a single place to which users must come and visit, there is need for digitised content to be dispersed on the Internet, passing through as many channels as possible. And beneath all this, there is still a need for sophisticated forms of organisation to underpin the delivery of digitised material. The lessons from earlier digitisation work still need to be fully digested and, more importantly, acted upon by all the relevant stakeholders. Creating the necessary structures, often in partnership, where institutions can build, deliver and curate their digital content will require a considerable amount of well organised planning and execution.

---

<sup>45</sup> To consider these issues, JISC has commissioned Chris Batt consulting to undertake a feasibility study on the digitisation and curation of collections from the general public. The outputs of this will be available at <http://www.jisc.ac.uk/whatwedo/programmes/digitisation/reports/digicuration.aspx> from Autumn 2009.