**JAMES — Beyond**

<0 images>

**Beyond the Library Walls: The National Library of Wales Research Programme in Digital Collections**

James, R. and McCann, P.

The National Library of Wales (NLW) has been an enthusiastic adopter of digitization and networked technologies since the late 1990s as a means of opening up the nation’s cultural heritage. The digital plays a key role in the Library’s core strategic aim of ‘Knowledge for All’, not only to provide greater access to resources online, but also to enhance our potential impact and help us engage more effectively with our users, both existing and future (NLW, 2014). To address the challenges of delivering effective, usable, and sustainable digital resources, the Library established its own Research Programme in Digital Collections in 2011 to contribute to the development of a more coherent digital agenda.1 Adding value and impact to the Library’s digital content and the institution as a whole lies at the heart of what the Programme is seeking to achieve. It can no longer be assumed that digital resources, once built, will be used (Warwick et al., 2008). Attempts have therefore been made to examine potential barriers to the adoption and use of such resources and to better understand issues surrounding their sustainability.2 It is within this context that the Programme operates.

The Programme’s main areas of focus include developing an understanding of the use, value, and impact of the Library’s existing digital content; identifying ways of enhancing this content for research, teaching, and community engagement; and developing new digital content that addresses specific research and educational needs. This work is being realised through collaboration with partner libraries, museums and archives, cultural heritage organisations, universities, and other key stakeholders that cross institutions, collections, and disciplinary traditions, both nationally and internationally, with the aim of bringing tools, methods, and content together to create digital resources that are seamless and open. This paper will examine the work currently ongoing at the Library as part of its digital programme and, in particular, its attempts to foster collaboration in the creation and development of improved environments that enable more effective use, re-use, and linking of its digital content.

**Historical Digital Texts**

The creation of digital text, and especially historical digital text, is hugely challenging. The issues posed by the original text, including issues such as legibility and physical condition, are often exacerbated by attempts to process and subsequently analyse these texts digitally and will inevitably have implications for research. Such issues are demonstrated by problems of accuracy with the underlying OCR in the Library’s *Welsh Newspapers Online* project3—a free, searchable digital archive of over 100 historic newspapers of Wales, both Welsh- and English-language, dating from 1804 to 1919 in the Library’s holdings—which are often caused by blemishes and imperfections on the pages of the original. As Tim Hitchcock has recently argued, underlying issues of the design and structure of digital texts can be hugely problematic for research and algorithm-based searching (2013). Re-use and analysis of such content must, therefore, incorporate these concerns. As such, engaging expert users and curators of the material in the creation of a digital resource and the development of discovery and analysis tools is key.

**Collaboration and Co-Creation**

A collaborative approach was used by the Library in Cymru1914.org,4 a JISC-funded project to digitize primary sources relating to the Welsh experience of the First World War and its impact on all aspects of Welsh life, language, and culture. The project has virtually unified fragmented and often difficult-to-access materials from the libraries, archives, museums, and special collections of Wales to form a consolidated digital collection of interest to researchers, students, and the public on life in Wales during this significant period of change. The project was developed as a process of co-creation: a collaboration between the Library; the libraries, archives, museums, and special collections of Wales; and researchers of the literature, history, language, and politics of the period. Their input was used to inform the selection of content, the development of the user interface, and the testing of functionality. This model of active collaboration reunites curatorial and scholarly roles, redressing concerns regarding the gradual segregation of these roles in the current academic and cultural heritage landscapes (Schnapp and Presner, 2009). Such approaches also work towards further embedding digital resources in the life cycle of scholarly research, thus increasing its value, impact, and sustainability in the long term.

The Library is also looking to explore new collaborative opportunities to enhance its existing digital content through the application of crowdsourcing approaches (Dafis et al., 2014). The Library’s first collaborative crowdsourcing project, Cymru1900Wales,5 a classification project to gather the place names of Wales from the georeferenced Ordnance Survey maps of 1900, has been an opportunity for the Library to experiment with crowdsourcing and enable greater engagement with new and remote audiences. These possibilities are being developed through a co-supervised PhD with the University of Wales to assess the feasibility of such approaches for the transcription of the Library’s online collection of Welsh wills and to build virtual communities around the content. There is already significant interest in the wills, especially in family and local history circles. Crowdsourcing is an opportunity for the Library to use this interest in a positive way to add value to its collections and the institution itself, and to engage the community in its work. Such an approach gives the Library the opportunity to build closer links with its users across the world and to encourage connections between academia and the public. It is a way of allowing the public to meaningfully experience and understand their pasts, which, in essence, is the cultural heritage sector’s primary objective (Owens, 2012).

**Analysis and Linking of Content**

Users are becoming more demanding in their expectations of what digital resources can deliver. For these demands to be met and to ensure maximum usage, particularly within scholarship, environments must be created that allow for content to be analysed in more effective and complex ways. This includes resources that facilitate both close and distant readings of texts, allowing for the exceptional to be studied alongside large-scale aggregations of data (Moretti, 2005). The Library has experimented with simple visualisation tools, such as n-grams, giving researchers the opportunity to analyse patterns and trends on larger scales than would be possible in a non-digital environment. The possibilities of linked data are also under consideration as a means of transcending collection boundaries. Enabling links to be more easily made between documents within and across collections in this way would remove the data from its Welsh silo and allow it to be more easily incorporated into a broader, more international context. This would see the Library becoming a truly global institution for and of the Welsh and other Celtic peoples.

In keeping with its commitment to providing free and open access, NLW is also in the early stages of facilitating more advanced analysis of its digital content by opening up some of its raw data for others to download and interrogate for their own research purposes. These datasets will come from some of the Library’s biggest collections, including the newspapers, and it will be available during 2015.6

**Conclusion**

This paper will demonstrate that cultural heritage organisations, such as the Library, are frequently spaces in which research and development in the digital humanities can be nurtured and encouraged, and where collaboration with colleagues within and beyond the humanities and the academy can take place. NLW projects, such as Cymru1914.org, have shown the value in building better connections and partnerships with libraries, archives, and museums, along with researchers and the public more generally. This has enabled the development of digital resources that allow for data to be used and reused, that encourage meaningful engagement with different user communities, and that permit digital outputs to be repurposed for new and unforeseen purposes. We must look to build collaborations that will facilitate the integration of tools and data, and the disassociation of text and data from its platform or delivery mechanism, thus ultimately liberating digital resources and scholarship, and ensuring that our work does more than merely replicate print culture digitally.

**Acknowledgement**

The authors would like to thank Professor Lorna M. Hughes for her assistance in writing this proposal.

**Notes**

1. NLW Research, www.llgc.org.uk/research.

2. University College London, *The LAIRAH Project: Log Analysis of Internet Resources in the Arts and Humanities*, http://www.ucl.ac.uk/infostudies/LAIRAH/. Oxford Internet Institute, University of Oxford. *TIDSR: Toolkit for the Impact of Digitised Scholarly Resources*, http://microsites.oii.ox.ac.uk/tidsr/welcome.

3. Welsh Newspapers Online, http://welshnewspapers.llgc.org.uk/en/home.

4. Cymru1914: The Welsh Experience of the First World War, http://cymru1914.org/.

5. Cymru1900Wales, http://www.cymru1900wales.org/.

6. The data, when available, will be accessible from <http://data.llgc.org.uk>.

**References**

**Dafis, L., Hughes, L. and James, R.** (2014). What’s Welsh for ‘Crowdsourcing’? Citizen Science and Community Engagement at the National Library of Wales. In Ridge, M. (ed.), *Crowdsourcing our Cultural Heritage*. Ashgate,pp. 139–59.

**Hitchcock, T.** (2013). Confronting the Digital: Or How Academic History Writing Lost the Plot. *Cultural and Social History,* **10**(1): 9–23, http://dx.doi.org/10.2752/147800413X13515292098070.

**Moretti, F.** (2005). *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso.

**NLW.** (2014). *Knowledge for All: National Library of Wales Strategic Plan 2014–2017*. http://www.llgc.org.uk/fileadmin/fileadmin/docs\_gwefan/amdanom\_ni/dogfennaeth\_gorfforaethol/corff\_strat\_KnowledgeforAll\_2014\_2017S.pdf.

**Owens, T.** (2012). Crowdsourcing Cultural Heritage: The Objectives Are Upside Down. http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/.

**Schnapp, J. and Presner, T.** (2009). *The Digital Humanities Manifesto 2.0*. http://www.humanitiesblast.com/manifesto/Manifesto\_V2.pd .

**Warwick, C., Terras, M., Huntington, P. and Pappa, N.** (2008). If You Build It, Will They Come? The LAIRAH Study: Quantifying the Use of Online Resources in the Arts and Humanities through Statistical Analysis of User Log Data. *Literary and Linguistic Computing,* **23**(1): 85–102, 10.1093/llc/fqm045.