

---

## New potentials in the digital archives: a participative inquiry into interdisciplinary collaboration in digital historical research at the Wellcome Trust

**Alex Green**

a.green@wellcome.ac.uk

Wellcome Trust, United Kingdom

**Lalita Kapish**

l.kapish@wellcome.ac.uk

Wellcome Trust, United Kingdom

**Hannah Walker**

h.walker@wellcome.ac.uk

Wellcome Trust, United Kingdom

---

This paper discusses an exploratory project where a group of university academics, software developers, designers and librarians spent a week analysing a broad selection of the Wellcome Library's digital collections with an aim to explore new ways of conducting collaborative digital history research, identifying and documenting barriers and successes and also pointing to gaps in institutional support infrastructures.

Over the past two decades, significant quantities of cultural heritage have been digitised. Internet Archive has digitised over 20 million items in partnership with libraries and collections around the world. Google has digitised over 25 million books as Google Books. Alongside this the emergent field of digital humanities has sought to take advantage of new opportunities afforded by unprecedented access to collections. Numerous commentators and researchers have, in the words of Hayles (2015), voiced the opinion that "if there is an area of general agreement, it is the transformative potential of digital humanities for the humanities and for academic discourse" (see also Ogilvie, 2016; Alves, 2014).

Wellcome Library, part of the Wellcome Trust, is one of the world's most significant collections relating

to health and medicine, with works ranging from posters and paintings to personal archives, printed books and packaging ephemera. Through its digitisation programmes, Wellcome is a major producer of digitised historical material and datasets, with over 40,000 digitised archives, nearly 100,000 digitised monographs and over 10,000 artworks, manuscripts, videos and reports. These have been made freely available under the most liberal licence possible, dependent on the copyright status of the material. In addition, IIF image services (IIF Consortium) including standardised image and presentation APIs, along with services for OCR and text search are provided (Chaplin, 2016).

A key aim for Wellcome is to enable new types of research and knowledge production; our mission is 'to improve health for everyone by helping great ideas to thrive' (Wellcome Trust, 2016). One way we seek to do this is through enabling the exploration of cultural and social meanings of health in the past. Digitisation has undoubtedly increased use of our collections significantly and developed a large international audience, with over 50% of researchers accessing content from outside the UK. To better understand the users and usage of our digital collections we have carried out quantitative and qualitative research in collaboration with Prof. Pauline Leonard (Green and Andersen 2016). Our findings showed that beyond increasing access, the nature of collection usage has not changed significantly. The majority of researchers still access works within a single collection, page through digitised works in a linear fashion, make limited use of OCR search and little to no use of APIs. We are not yet seeing Hayle's "transformative potential" (2015) realised. To better understand this, we developed a number of hypotheses and research questions, which we categorise here under four loose umbrellas:

### Lean working

In summarising the findings of multiple digital historical studies, Alves has argued that they 'implicitly confirm the efficiency of digital means... but also... that their application is, often, generally associated with expensive projects requiring extensive human resources with diverse skills' (2014). We questioned whether digital research is necessarily expensive or requiring of extensive resource. Drawing inspiration from commercial software development, we asked what is the Minimum Viable Product (Leanstack, 2016); can teams achieve meaningful research relatively inexpensively through an agile approach of iter-

ative investigation and identification of emergent areas of interest rather than pre-defining fixed research questions.

### Skills and knowledge

It could be argued that there is a skills or knowledge gap within traditional historical research communities which inhibits conducting digital research. However, we questioned if this was truly a barrier when evidence from citation patterns shows increases in collaborative approaches to digital humanities research (Nyhan and Williams, 2013). We questioned if extending the scope of traditional research teams to include commercial development partners could bring new insights and capabilities.

### Crossing collections

Wellcome's digital collections include intensely heterogeneous material along with data sourced from multiple institutions. We questioned whether there are practical or technical barriers to break down divisions between collections and drawing from a range of content. Hitchcock (2013) has argued that 'the lack of flexibility of the available digital tools [has] enabled only the effective utilization and analysis of quantitative sources or sources easily transformed into a quantitative format'. As many of Wellcome's collections are archival and contain large quantities of handwritten and pictorial material, we specifically wanted to explore the possibilities of digital for non-quantitative research, and research which still requires 'close reading' of sources (Van Dijk, 1985).

### Quality and consistency

Areas of enquiry in this umbrella included whether our digital collection is suitable in its scale and scope, and the quality and consistency of OCR and collection metadata. We also questioned if we had the right kind of web services available and their usability for individuals with different levels of experience.

To explore and better understand these questions we designed an experimental approach, adapting a concept becoming familiar to cultural institutions – the 'hack day' – and extending it out into a week-long intensive R&D project, where small teams led by a mix of independent and academic researchers would work collaboratively to explore the Wellcome Library digital collections. Our participants included research staff from several UK universities, librarians and archivists, commercial designers and software developers. Through careful screening of participants, we selected researchers with shared enthusiasm for, but variable

experience of digital historical research. This choice was deliberate in order to focus in on barriers relating to experience, skills and technical feasibility without confounding these variables with any reluctance to use digital methods. However, the enthusiasm for digital methods in the historical community is an important question and undoubtedly merits further investigation. Research areas were open-ended, with a focus on experimentation rather than production of finished work. However, we did agree broad areas, including handwritten records from a private asylum, 5000 Medical Officer of Health reports covering London from 1848-1972, 6,600 issues of the trade journal *Chemist and Druggist* and the 79,000 books digitised by the UK-MHL project (Wellcome Library, 2016).

Drawing from approaches of participative and co-operative inquiry (Reason and Bradbury, 2008) we embedded Wellcome staff in mixed teams of academic staff, developers and designers, positioning the teams as researchers of the collections, participants in a broader experiment of research production and observers and documenters of the experiment itself. We encouraged self-documentation by teams through use of project boards, blogs and wikis, conducted individual interviews with participants throughout the process and held daily reviews and a plenary session to discuss progress and reflect on the experiment.

During the week we produced multiple tools and visualisations, and also historical findings. These are documented on the [project blog](#), along with the processes each team went through. This clearly demonstrated the potential of the collections, but also the barriers to working with them. Interestingly, the project using digitised archival material without OCR was one of the most successful – partly due to the synergy of the team members, but also the clarity of the challenge for the material. Two of the five projects seeded in the week continue to be developed, and we are continuing to provide support to the researchers leading them.

Feedback from participants in the project identified particularly the benefits of working in teams with mixed skillsets. Developers and library staff gained from acquiring greater understanding of research processes and interests, while researchers gained access to technical skills, and exposure to a different approach to digital working. One participant remarked: 'I found working with other people, from a variety of different backgrounds, really generative – both for thinking about why people who do different kinds of work approach digital resources in different ways, and for thinking about my own research along new lines.'

The web services we provide for programmatic access to digital collections were identified as a particular barrier during the week. While the images and bibliographic metadata for our collections are exposed through an international standard, the Image Interoperability Framework (IIIF), this proved conceptually complex for developers to quickly prototype with, requiring combining and processing multiple JSON responses to access digital items. Adding further complexity to this, our collection OCR is available primarily through ALTO XML (Library of Congress, 2016) which our developers found challenging to process at scale.

As cultural heritage institution and a research funder, we are continuing to unpack the implications of the findings. As a library, we found great value in co-production with research users so will be repeating a similar annual event, investigating different aspects of our collections. We have also identified particular issues related to the usability of our collections and added these to our development roadmap. As a funder we are considering options for increasing innovation by seeding early stage research through similar collaborative processes. To take this forward we will be running a series of pilot events through our interdisciplinary research residency, [The Hub](#).

## Bibliography

- Alves, D.** (2014). Guest Editor's Introduction: Digital Methods and Tools for Historical Research. *International Journal of Humanities and Arts Computing*. **8**(1):1-12.
- Bergold, J. and Thomas, S.** (2012). Participatory Research Methods: A Methodological Approach in Motion. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*. **13** (1): Art. 30, <http://nbn-resolving.de/urn:nbn:de:0114-fqs1201302>.
- Chaplin, S.** (2015). Why Creating a Digital Library for the History of Medicine is Harder than You'd Think! *Medical History*. **60**(01):126-129.
- Green, A & Andersen, A.** (2016). 'Finding value beyond the dashboard', paper presented to MW2016: Museums and the Web 2016, Los Angeles, 6-9 April. Available at <http://mw2016.museumsandtheweb.com/proposal/guaging-value-beyond-the-dashboard/> (Accessed: 1 November 2016).
- Hayles, N.** (2015) How We Think: Transforming Power and Digital Technologies. In Svensson, P. and Goldberg, D. (eds), *Between Humanities and the Digital*. MIT; 2015. p. 503.
- Hitchcock, T.** (2013). Confronting the Digital. *Cultural and Social History*. **10**(1):12-14.
- IIIF Consortium.** (no date). *International image Interoperability framework*. Available at: <http://iiif.io/about/> (Accessed: 1 November 2016).
- Library of Congress.** (2016). *About ALTO*. Available at: <http://www.loc.gov/standards/alto/about.html> (Accessed: 5 April 2017)
- Leanstack.** (2016). *Minimum viable product*. Available at: <https://leanstack.com/minimum-viable-product/> (Accessed: 1 November 2016).
- Ogilvie, B.** (2016). Scientific Archives in the Age of Digitization. *Isis*. **107**(1): 77-85.
- Nyhan, J. and Duke-Williams, O.** (2016). Joint and multi-authored publication patterns in the Digital Humanities. *Arche Logos*. <http://archelogos.hypotheses.org/103>
- Reason, P. & Bradbury, H.** (2008). Introduction. In Reason, P. and Bradbury, H.(eds), *The Sage handbook of action research. Participative inquiry and practice*. London: Sage, pp.1-10.
- Toon, E, Timmermann, C & Worboys, M.** (2016). Text-Mining and the History of Medicine: Big Data, Big Questions?. *Medical History*. **60**(02):294-296
- van Dijk, T.A.** (ed.) (1985). *Handbook of discourse analysis: V. 1: Disciplines of discourse*. 3rd ed. Orlando: Academic Press.
- Wellcome Library.** (2016). *UK medical heritage library*. Available at: <https://wellcomelibrary.org/collections/digital-collections/uk-medical-heritage-library/> (Accessed: 1 November 2016).
- Wellcome Trust.** (2016). *Our Strategy*. Available at: <https://wellcome.ac.uk/about-us/our-strategy> (Accessed: 1 November 2016).