

## Interface development, including bi-lingual support

Users appreciate straightforward user interfaces, so a simple bilingual interface was developed to provide material in a variety of formats and at varying levels of archival complexity, while retaining the hierarchical structures of archives that increase usability – through familiarity – of digital resources.

An interface group conducted on-going, iterative usability testing and implementation, including several user workshops: a formative evaluation exercise; an education workshop; and a participatory design workshop, organized by the Humanities Research Institute at the University of Sheffield, who are working on a project entitled Participating in *Search Design: a study of George Thomason's English Newsbooks* (<http://>). The goal of the latter workshop was to see if the participatory design methodology could feed into development of the interface by engaging with potential end users[5]. This paper will supply data about the above activities and they will be presented with cross referral against specific user communities. As the resource is used information gathered will be used to generate user case studies.

## Dissemination and stakeholder engagement

Stakeholder engagement throughout the development of the project was crucial to ensuring widest use and re-use of the content, via a process of collaboration and outreach to disparate user communities, and usability testing and engagement with the digital outputs of the project. The project team worked with core communities through an iterative process of engagement and input throughout the development of the project, through the establishment of a research network of academics using the content, specifically participating in the three stakeholder workshops, and five Community engagement workshops, organised by the People's Collection Wales (<http://www.peoplescollectionwales.co.uk>). Post-launch user data from products such as Google Analytics also shape our findings.

## Sustainability

In many respects, the actions described above to promote use, uptake and embedding of the resource are the surest way to ensure sustainability: digital collections that are used will be sustained over the long-term as they become invaluable to education, research, and community building.[6] A recent report by ITHAKA for the Strategic Content Alliance, "Sustaining our Digital Future"[7] highlighted the need to make planning for sustainability a key component of the digital life cycle. The use of good practice in digitization, and the use of an open-source, scalable repository such as Fedora, is key to sustaining the digital objects, of course, but key to cultivating sustainability of our valuable digital content is to embed planning for impact into the planning and development of digital resources. Fedora is a vital component of our long-term sustainability plans, and our institutional setting is key to this. Providing a crucial resource for research, teaching and public engagement around the topic of the Welsh experience of the First World War will promote sustainability of the resource. A key factor in planning and designing the resource as described in this paper is to create a digital content platform that can be added to over time. We also plan to revisit the use of the resource and to use this summative evaluation as the basis for any required modifications to increase its use.

## Conclusion

It is increasingly obvious that factoring in end use of digital resources as broadly as possible at the outset of a digitization project is crucial: impact is a crucial component of the entire digital life cycle. The ultimate use of digital materials is a consideration that impacts decisions made at every stage of this life cycle: selection, digitization, curation, preservation, and, most importantly, sustainability over the long term. The way that digital resources are used may be unanticipated at

the outset; or they may have value for different communities and disciplines than originally intended. The best resources have been developed in such a way that their use and re-use has been anticipated at the outset, and that unforeseen use is anticipated through the use of technical standards and approaches. Just as digital collections that have been developed in formats that are not "open" are far less likely to be re-used for teaching or research, if digitization is to have more impact than being a form of "digital photocopying", the user needs to be placed at the centre of the process from the outset.

## References

- Hughes, L.M.** (2013) "Digital Collections in the Humanities: Understanding Use, Value and Impact", special issue of 'Digital Studies / Le champ numérique' from SSHRC Montreal seminar on cyberinfrastructure
- Warwick, C., Terras, M., Huntington, P., Pappa, N., & Galina, I.** (2006). *The LAIRAH Project: Log Analysis of Digital Resources in the Arts and Humanities Final Report to the Arts and Humanities Research Council*. London: University College London. Available online: <http://www.ucl.ac.uk/infostudies/claire-warwick/publications/LAIRAHreport.pdf>.
- Hughes, L.M., Eil, P., Dobрева, M., and Knight, Gareth, K.** forthcoming (2013) "Assessing and measuring impact of a digital collection in the humanities: An analysis of the SPHERE (Stormont Parliamentary Hansards: Embedded in Research and Education) Project", *Literary and Linguistic Computing* (Oxford)
- See Dobрева, M., O'Dwyer, A., and Konstantelos, L., in Hughes, L.M.** (ed) (2011) *Digital Collections: Use, Value and Impact*. London: Facet
- For an overview of this approach, see: **Wessels, B., Dittrich, Y., Ekelin, A and Eriksen, S.** (2012). 'Creating synergies between participatory design of e-services and collaborative planning' in *International Journal of E-planning Research*, Volume 1, Issue 3, doi: 10.4018/ijep.2012070101
- Hughes, L.M.**, (2011) "ICT Methods for digital collections research, chapter in Hughes, L.M. (ed) (2011) *Digital Collections: Use, Value and Impact*. London: Facet
- <http://www.sr.ithaka.org/research-publications/sustaining-our-digital-future>

## Using digitized newspaper archives to investigate identity formation in long-term public discourse

**Huistra, Hieke**

*h.m.huistra@uu.nl*  
Utrecht University

**Pieters, Toine**

*t.pieters@uu.nl*  
Utrecht University

This paper analyzes how digitized newspaper databases can be used in historical research on identity formation in public discourse. It discusses a new semantic text mining tool, Texcavator, which is currently being developed in the Dutch research program *Translantis: Digital Humanities Approaches to Reference Cultures*.<sup>1</sup> The paper presents a case study which combines the Texcavator tool with the publicly available Delpher<sup>2</sup> and with traditional historical methods in order to analyze identity formation of health risk groups in Dutch public discourse in the twentieth century. In particular, it focuses on the construction of the identity of people with excess body weight. Although the case is built around the Texcavator and Delpher mining tools and the newspaper database of the Dutch national library, the paper aims to investigate techniques to combine close and distant reading that can be transferred to other tools and repositories as well.

Newspapers are valuable sources in historical research. Until recently, however, investigating them was cumbersome

and time-intensive. The repositories of digitized newspapers now available in many countries solve many practical problems and offer wonderful opportunities, but they also introduce methodological problems of their own. (Bingham 2010; Nicholson 2013) Bob Nicholson has recently shown how digitalization enables us to approach newspapers bottom-up instead of top-down, but he stresses the difficulty of creating useful keyword searches for doing this. (Nicholson 2013, 66–67) Adrian Bingham has also pointed this out, and has furthermore highlighted the danger that keyword searches (as well as other text mining techniques) pluck individual articles out of their original context, ignoring their position on the page, surrounding articles, and illustrations. (Bingham 2010, 230) Furthermore, Johanna Drucker has indicated that digital humanities scholars often aim to reduce complexity and remove ambiguity, while these are two values humanities research has to cherish, not avoid. (Drucker 2009, 5–7; Collini 2012, 65–84)

This paper takes such warnings into account and shows how these problems are being addressed by researchers working with the digitized newspaper database of the Dutch national library, thereby offering more concrete versions of the rather general solutions (e.g., ‘we should not forget the article’s context’) that are often suggested. At present, this database contains over 10 million pages from more than 200 newspapers and periodicals published between 1618 and 1995.<sup>3</sup> It can be approached in two ways: through Texcavator (in development, not yet publicly available) and through the national library’s Delpher tool (publicly available).

The paper discusses a specific use case in which both tools are combined and used alongside traditional historical methods: researching identity formation in public discourse. It focuses on the identities of (health) risk groups, groups of people that are classified as ‘at risk’ with help of (health) risk factor classifications like the body mass index (BMI). For example, nowadays, people with a BMI above 25 are classified as ‘at risk’ because of their high body weight. This classification and the construction of this group is not a necessary outcome of biomedical research on the human body; instead it is historically contingent, strongly rooted in culture and practice. (Hacking 2007a, 2007b) The construction of these risk groups and the formation of their identity takes place for a significant part in public discourse. Digitized newspapers are valuable sources to study this identity formation: they provide a good entry into public discourse and typically span long time periods, enabling researchers to analyze the fluctuations in the identity of these groups (e.g., fluctuations between whether or not they are seen as (and see themselves as) ‘ill’).

The paper presents the first results of the investigation of the identity construction of the risk group ‘overweight people’ between 1890 and 1990. It focuses in particular on newspaper advertisements in the first part of this period — a choice based on distant reading of the corpus with help of Texcavator. The paper discusses how Texcavator and Delpher have been used, focusing in particular on the interaction between close and distant reading necessary to do this type of research. It shows how the direct connection between Texcavator and Delpher makes sure the researcher is constantly only one or two mouse clicks away from viewing the single articles in their original context — on the page, including illustrations, within the full issue of the periodical, as if going through newspapers on microfilm (or, depending on the size of the computer screen, leaving through them on broadsheet). Furthermore, it shows how Texcavator’s built-in visualization tools (time lines with number of articles diagrams, word clouds, named entity recognition) can be used to go back and forth between distant and close reading in order to build sophisticated queries that can easily be refined and modified within the tool.

In this way, the paper shows the challenges but also the new heuristic possibilities of doing historical research in digital repositories of newspapers.

## Notes

1. [www.translantis.nl](http://www.translantis.nl)

2. On Texcavator see **Huijnen et al.** 2013; on Delpher see [www.delpher.nl](http://www.delpher.nl). Delpher has replaced the earlier, no longer available tool Lucene.  
3. [delpher.kb.nl](http://delpher.kb.nl)

## References

- Bingham, A.** (2010). *The Digitization of Newspaper Archives: Opportunities and Challenges for Historians*, Twentieth Century British History, 21: 225–231.
- Collini, S.** (2012). *What Are Universities For?* London: Penguin.
- Drucker, J.** (2009). *SpecLab: Digital Aesthetics and Projects in Speculative Computing*. Chicago: University of Chicago Press. Available at <<http://public.eblib.com/EBLPublic/PublicView.do?ptilID=448540>>.
- Hacking, I.** (2007a). Kinds of People: Moving Targets. *Proceedings of the British Academy*, 151: 285–318.
- Hacking, I.** (2007b). Where Did the BMI Come From. In: *Bodies of Evidence: Fat Across Disciplines*. Cambridge: Centre for Research in the Arts, Social Sciences and Humanities, University of Cambridge.
- Huijnen, P., Laan, F., de Rijke, M., Pieters, T.** (2013). A Digital Humanities Approach to the History of Science: Eugenics Revisited in Hidden Debates by means of Semantic Text Mining. In: Wierzbicki, A., Jatowt, A., Nadamoto, A., Leidner, J. (eds.), *SocInfo 2013 Workshop Proceedings, 1st International Workshop on Histoinformatics*, Cham et al: Springer.
- Nicholson, B.** (2013). The Digital Turn. *Media History*, 19: 59–73.

## Extracting Relationships from an Online Digital Archive about Post-War Queensland Architecture

### Hunter, Jane

The University of Queensland, Australia

### Macarthur, John

The University of Queensland, Australia

### Van der Plaats, Deborah

The University of Queensland, Australia

### Gosseye, Janina

The University of Queensland, Australia

### Muys, Andrae

The University of Queensland, Australia

### Macnamara, Craig

The University of Queensland, Australia

### Bannerman, Gavin

The State Library of Queensland

The “Architectural Practice in Post-War Queensland: Building and Interpreting an Oral History Archive” project is a collaboration between the University of Queensland, the State Library of Queensland (SLQ) and four of the longest-standing architectural firms in Queensland. The project’s aim is to build a comprehensive online multimedia digital archive that documents architectural practice in post-war Queensland (1945–1975) — a period that was highly significant in Queensland’s architectural history but that remains largely undocumented. The goal was to use innovative Semantic Web technologies to link tacit knowledge extracted from individual oral histories to tangible knowledge (drawings, books, photographs, manuscripts) that exists within personal archives, firm archives as well as State and institutional archives and libraries.

The approach involved firstly conducting and recording a series of oral history interviews and public forums with the key architects from this period. These events comprise both private