
We're All Experts Now? Archiving public health discourse in the UK Web Archive

Austin Alice*^{†1}, Cui Cui*^{‡2}, and Leontien Talboom*^{§3}

¹University of Edinburgh – United Kingdom

²Bodleian Libraries – United Kingdom

³Cambridge University Library – United Kingdom

Abstract

The UK Web Archive (UKWA) has been preserving UK websites since around 2013, but the COVID pandemic catalysed a rising crisis for archiving rapidly changing information shared on the web. Emerging from these COVID-19 collecting initiatives, the Archive of Tomorrow is an ambitious collaborative project between the National Library of Scotland, Edinburgh University Library, Cambridge University Library, and Bodleian Libraries, Oxford that intends to curate a collection of c.10,000 health-related sites.

This collection broadens the scope of previous Covid-related collecting to include wider public health issues and health information – its accuracy, context, communication, and impact. Unlike other collecting initiatives that have targeted particular types of websites (e.g. open government websites), the project's approach will capture a more representative and diverse collection of public health websites in the UK. Scoping information from dissenting or contentious actors as well as from official and authorised sources will support future research into the interaction between health and the internet, and it is anticipated that the collection will be used as a test bed with which to interrogate the creation, management, and use of archived web resources for research.

The project has encountered a number of challenges: there are technical barriers to be overcome in capturing interactive and dynamic sites, ethical considerations to be made concerning how disputed or outdated information might be responsibly made available to researchers, and philosophical questions to be asked about how 'health information' is to be defined.

This paper will report on the outcomes of the project (which will be wrapped up in April 2023) and discuss future directions for improving the production and use of large-scale archived web collections.

Keywords: health information, web archiving, access

*Speaker

[†]Corresponding author: alice.austin@ed.ac.uk

[‡]Corresponding author: cui.cui@bodleian.ox.ac.uk

[§]Corresponding author: lkt39@cam.ac.uk