

Preserving born digital conspiracy theories and online emotions: webarchiving the Dutch national domain in the Post-truth era (abstract)

Kees Teszelszky

According to the Oxford Dictionaries, Post-truth is relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief. These circumstances became especially present on the web in our time, as hate speech, extreme political beliefs and alternative facts are flooding web fora and social media. The Leiden academic Jelle van Buuren has done a PhD research on conspiracy theories on the web (2016). He addressed the question whether or not conspiracy constructions are boosting hatred against the political system. The online discourse in which these conspiracy theories get shape can be characterized as Post-truth.

This paper will describe the attempts of The Koninklijke Bibliotheek – National Library of the Netherlands (KB-NL) to construct a web collection which reflects the Post-truth era. It started with a link extraction of the dissertation of Van Buuren. This collection contains a selection of relevant sites from the Dutch national web domain. In this paper, I will outline the use of this web collection and relevant datasets for current and future digital humanities research. It will be argued that the process of web archiving must not be limited to selecting and harvesting sites, but also to map and preserve the digital context of this source material from the web in datasets. These sources are a combination of various interlinked datasets, which together prove the authenticity of the source material.

KB-NL as a national library has collected born digital material from the web since 2007 through web archiving. It makes a selection of websites with cultural and academic content from the Dutch national web. Most of the sites were harvested because of their value as cultural heritage. Due to these selection criteria, many of the harvested websites are about history and heritage. As such, they are less suitable to use as primary sources for future historic research on the our own digital age.

I will describe the methods and experience with selecting, harvesting and archiving websites of the Post-truth era. I discuss also the characteristics of web materials and archived web materials and will explain the use of these various materials (harvested websites, link clouds, context information) for digital humanities research. I will also argue that the combination of a growing variety of different kinds of digital materials and processes from the web calls for a reinterpretation of primary historic sources, stressing the question what can be regarded as an authentic born digital source of our time.