Leveraging mailing list archives for digital history research: A case study on mining the Humanist discussion group

Marco Humbel[1,2], Jiajie Zhang[1], Valentina Prishchepova[2], Aleksey Varfolomeyev[2], Julianne Nyhan[1,2] and Andreas Vlachidis[1]

[1] University College London, United Kingdom

[2] Technische Universität Darmstadt, Germany

What can mining of mailing lists tell us about the formation of academic disciplines? How can computational tools help to uncover geographical, technical, disciplinary and institutional dynamics present in mailing list correspondences? And what are the challenges in leveraging mailing list archives as primary source material for historical scholarship? We will respond to these questions in context of the Mixed Methods Digital Oral History (MeDoraH) project in mining 38 years of the Digital Humanities mailing list discussion group the 'Humanist'. MeDoraH takes the history of the Digital Humanities discipline (ca. 1949 to the present day) as a case study to explore the potentials of combining oral history research with digital research methods and source material, such as the Humanist's mailing list archives.

The Humanist was established in 1987 as "a Bitnet/NetNorth electronic mail network for people who support computing in the humanities" (McCarty, 1987). With over 40,000 contributions in 38 years the Humanist presents a unique source to deepen our understanding of the dynamics that shaped the formation of the Digital Humanities discipline. The Humanist is also an excellent example of the increasing relevance of born digital primary source material for historical research. This is because it is impossible to understand the history of the 21st century without taking born digital sources into account (Owens and Padilla, 2021: 329). Yet, according to Brügger sources like the Humanist are considered to be sitting "on the edge of the web". The histories recorded in mailing list archives tends to be unwritten as there is a lack of preservation, public access and accordingly lack of experience of dealing with them as source material (2018: 149–51).

In this presentation we will discuss the affordances posed by mailing list archives for digital history research in terms of data harvesting, preparation and ethics for data access and use (UCL ethics ID: AH/2025/23). In our explorative study we extracted named entities with spaCy and conducted LDA (Latent Dirichlet Allocation) topic modelling with Mallet. Through this approach we identified patterns of people and companies' influence on the formation of the Digital Humanities. By comparing the results with networks of recollection in the Hidden Histories oral history corpus (Nyhan et al., 2015) we will reflect on the potentials and challenges of mailing list archives for moving towards a 'digital methodological-hermeneutical turn' in oral history research. Accordingly, our work advances the establishment of best practices for leveraging mailing list archives in digital scholarship.

References

Brügger, N. (2018). *The Archived Web: Doing History in the Digital Age*. Cambridge, MA: The MIT Press.

McCarty, W. (1987). https://dhhumanist.org/archives/Converted_Text/humanist.1987-1988.txt (accessed 25 April 2025).

Nyhan, J., Flinn, A. and Welsh, A. (2015). Oral History and the Hidden Histories project: towards histories of computing in the humanities. *Digital Scholarship in the Humanities*, **30**(1): 71–85 doi:10.1093/llc/fqt044.

Owens, T. and Padilla, T. (2021). Digital sources and digital archives: historical evidence in the digital age. *International Journal of Digital Humanities*, **1**(3): 325–41 doi:10.1007/s42803-020-00028-7.