

## **DH Benelux 2024 – Breaking Silos, Connecting Data: Advancing Integration and Collaboration in Digital Humanities.**

### **Sharing is caring. Archiving social media together (short presentation)**

Katrien Weyns, KADOC-KU Leuven

Data silos, that is what social media platforms manage, and it has been the premise of their revenue model for years. As the data-driven economy gained momentum, we saw these platforms increasingly closing the doors of these silos. We are on a point where they only open them on request and access is granted under very strict conditions. It is becoming increasingly difficult or even impossible to extract data from widely used platforms such as Facebook and X (Twitter) for research or for long-term preservation as cultural heritage. However, the impact of social media on society is too great to ignore. Not preserving or researching these data streams would lead to a substantial gap in our legacy and the influence of the social media platforms themselves.

GLAM (galleries, libraries, archives and museums) in Belgium, which increasingly recognize the cultural heritage value of this medium, lacked the expertise to build social media archives until recently. Efforts were mostly limited to short-term experiments because the threshold to work structurally on social media archiving is very high due to the variety of platforms, the complex and dynamic nature of the medium, and the lack of a clear legal framework. Between 2020 and 2023, KADOC-KU Leuven, archive and documentation center for heritage on the social embedding of religion in Belgium after 1750, therefore examined best practices for social media archiving by cultural archives together with 11 other GLAM partners. During the project KADOC brought together experiences with capture tools, archiving models were tested, the legal framework was explored, and access and reuse options were tested in pilot projects. The research was complemented by experiences from the BESOCIAL project (KBR), similar experiments of Dutch heritage institutions and RDDI (Rijksprogramma voor Duurzaam Digitale Informatiehuishouding of the Dutch government), which conducted concurrent research and experiments. Also projects of IIPC, International Internet Preservation Consortium, contributed insights.

It soon became clear that we are facing similar problems in Belgium and in Europe, and that this the case for both small and large institutions. The main problem is increasingly about getting access to data - will your request be granted - in addition to the technical aspect - how to collect the volatile data itself and what precisely you can collect. These challenges are similar for researchers, who want to get started with social media for quantitative as well as for qualitative research, as they are for GLAM in the Low Countries: how do we keep up with technical-organizational evolutions? How do we obtain and maintain legitimate access to data? How do we enable analysis and (re)use?

It is therefore necessary to build bridges between both disciplines and what better way to do so than during DH Benelux. By connecting or better engaging the academic and GLAM initiatives, we can better address the common challenges that exist. There are plenty of opportunities. First, there is in the area of expertise sharing. Available crawling and scraping tools change frequently and are often platform-dependent. Choosing the most appropriate tool can be a time-consuming process. It saves time if you can fall back on existing networks and experiences of others. Maintaining easily accessible manuals for tools is also a form of expertise sharing and building. Second, cooperation can be situated in the area of preservation and making archived data available. Cultural heritage institutions have the task of preserving cultural heritage in the long term for its cultural-historical value and making it available (in many cases this is done conditionally). They facilitate research, but are also open to include data streams from researchers if they fit into their collection. This keeps the data available for reuse. Do researchers know of these services the cultural heritage sector can provide? Third, it is important to work interdisciplinary to raise awareness and communicate about our needs to make social media archiving and research possible at various policy levels. Especially in the area of social media access and preservation, it is desirable that voices reinforce each other.

In the presentation, the results of the research on best practices for social media archiving in a GLAM setting will be the starting point to invite researchers to get started with the published manuals and overviews. The project also led to the creation of a practitioners network of actors with a lot of no experience that twice a year shares experiences and updates on social media archiving and research in lectures, workshops and panels. In this network, the proportion of researchers is still too limited. A presentation of project results should also be an invitation for DH researchers to engage in conversation with the heritage sector. Archival institutions mostly keep captioned social media behind the scenes, where they are available for research reuse. So it pays for researchers to knock on doors to get a better idea of the resources available. Shared efforts can also be a slant to share workloads on data collection, which can be very labor-intensive for qualitative research, and create a win-win.