

## **Navigating legal concerns in social media archiving**

**Social media has radically changed the way we communicate and document events, enabling anyone with access to the internet to create and share ideas, memories and perspectives on a variety of topics on an unprecedented scale. It is because of the enduring cultural value such material holds that cultural heritage institutions are increasingly seeking to include it in their collections, despite the many challenges.**

Developing a social media archive, however, requires first and foremost a careful consideration of the legal and ethical landscape within which such collections are created, managed, and made accessible. This means considering and assessing regulatory aspects, including copyright, data protection laws and terms of service of social media platforms.

## **Key legal concerns to consider when collecting social media**

### **Ownership of Data and Copyright Concerns**

Ascertaining the ownership of content shared on social media, even with relative certainty, can be extremely challenging due to the stratification of stakeholders who may generate and share content, including attached media such as videos, images, and audio recordings. It is also important to note that, when signing up to platforms, users often pay little attention to terms of service without fully considering their implications. Many of these agreements state that, although users retain ownership of the content they create, by posting or submitting material they grant platforms, such as X.com, a “non-exclusive, royalty-free licence [...] to use, copy, reproduce, [...] display and distribute such content in any and all media or distribution methods now known or later developed.”<sup>1</sup> Such provisions add an additional layer of complexity to questions of ownership of data and raise significant ethical concerns, particularly given that users may be unaware of how their content can be used by third parties.

Moreover, archivists should consider that intellectual property rights may apply to social media on a two-pronged level: the creator who generates the content and the company who owns, for example, the design of the interface and logo. As web archiving practices consist of producing copies of websites which are then replayed via specific access interfaces,<sup>2</sup> this may raise concerns when the act of web archiving involves copyrighted material for which archiving initiatives may have not received explicit authorisation from the author. In some countries, institutions may benefit from specific copyright exemptions allowing them to capture content for preservation purposes.<sup>3</sup> However, such exemptions often come with restriction on modalities of access, (re)use of data and reproduction of web archived material, which may hinder engagement with web and social media collections.<sup>4</sup>

## **Data Protection Laws**

For archives and libraries “archiving in the public interest”, Article 89 of the GDPR provides specific exemptions. However, appropriate safeguards must be put in place to protect the data subjects’ rights, such as anonymization and minimization. Nevertheless, due to the sheer volume of data generated on social media platforms, even when applying all due precautions at the selection stage, it is important to take into account that some sensitive or personal information may inadvertently be captured when archiving at scale.

## **Social media policies**

Social media policies impose a set of limitations on collection activities in order to protect primarily their legitimate commercial interests. These include, for example, restrictions on the amount of data that can be accessed in a specific timeframe via platforms’ official APIs, as well as systems designed to detect behaviour patterns that could be attributed to web crawlers. While web scraping is somewhat legal, the risk of infringing platforms Terms & Conditions may still persist when not explicitly allowed by the platform itself.<sup>5</sup>

In addition, as these policies tend to change rather rapidly and unexpectedly, collecting institutions need to continuously adapt and develop new approaches to effectively collect content from these platforms.

## **Conclusions: recommendations and strategies to mitigate concerns**

In light of the, albeit brief, overview of challenges outlined above, social media archiving initiatives must consider adopting transparent and well-documented approaches to social media collection development and access. Every decision taken in response to legal constraints and ethical concerns should be clearly recorded and made available to users of the resulting collections, enabling them to critically engage with the material.

To mitigate privacy or copyright concerns, for example, permission-based approaches have proved to be a widely implemented strategy across community of practice. However, the time-consuming process of seeking permission to archive social media content is only feasible for small scale archiving projects. The sheer volume of data generated daily on social media, combined with its interconnected nature, makes it impractical to seek permission at scale, particularly given the limited resources available to many cultural heritage institutions to effectively support such endeavour.

Where possible, institutions should assess archived content for the presence of sensitive or personal information upon collection/donation. However, given the scale of many social media collections, item-level review may not always be feasible. In such cases, good practice includes the use of collection-level disclaimers to warn users about the potential presence of sensitive material. Limiting access to social media collections to vetted or accredited researchers, can offer an additional layer of risk mitigation.

Finally, following the example set by several web archiving institution, implementing a well-defined takedown procedure, including for example a list of clear criteria under which content may be removed from access (ideally developed in collaboration with in-house legal teams or under external advice), can help ensure transparency and further mitigate ethical and legal concerns.

<sup>1</sup> X.com, 'X Terms of Service', X.Com, November 2024

<<https://web.archive.org/web/20251009013025/https://x.com/en/tos>> [accessed 28 October 2025].

<sup>2</sup> Niels Brügger, *The Archived Web: Doing History in the Digital Age* (MIT Press, 2018), p. 80.

<sup>3</sup> S. Chambers and others, *BESOCIAL: Final Report WorkPackage1 an International Review of Social Media Archiving Initiatives.*, April 2021.

<sup>4</sup> See for example the report: Sharon Healy and others, *Skills, Tools, and Knowledge Ecologies in Web Archive Research*, published online 2022, doi:DOI%2010.17605/OSF.IO/VF7GT.

<sup>5</sup> Beatrice Cannelli, 'Archiving Social Media: A Comparative Study of the Practices, Obstacles, and Opportunities Related to the Development of Social Media Archives', in *Digital Humanities Research Hub* (unpublished doctoral, School of Advanced Study, 2024), p. 192 <<https://sas-space.sas.ac.uk/10023/>> [accessed 7 January 2025].



## **Beatrice Cannelli**

Researcher in Born-digital Cultural Heritage, Digital Humanities  
Research Hub, School of Advanced Study, University of London.

### **Abstract**

#### **English**

Cultural heritage institutions are increasingly engaging in the preservation of social media content. Yet, the complexity of social platforms requires careful consideration of the legal and ethical landscape before initiating any archiving activities. Concerns surrounding copyright, data ownership, data protection laws, and platform policies significantly shape social media collections. Documenting decision-making processes and implementing clear policies to address and mitigate these issues are therefore essential.

#### **Deutsch**

Kulturelle Einrichtungen beschäftigen sich zunehmend mit der Aufbewahrung von Inhalten aus sozialen Medien. Die Komplexität sozialer Plattformen erfordert jedoch eine sorgfältige Abwägung der rechtlichen und ethischen Rahmenbedingungen, bevor Archivierungsmaßnahmen eingeleitet werden. Fragen rund um Urheberrecht, Dateneigentum, Datenschutzgesetze und Plattformrichtlinien prägen die Gestaltung von Sammlungen aus sozialen Medien maßgeblich. Daher ist es unerlässlich, Entscheidungsprozesse zu dokumentieren und klare Richtlinien zu implementieren, um diese Probleme anzugehen und zu mindern.

#### **Français**

Les institutions chargées du patrimoine culturel s'engagent de plus en plus dans la préservation des contenus issus des réseaux sociaux. Cependant, la complexité de ces plateformes exige d'examiner attentivement le contexte juridique et éthique avant de lancer toute activité d'archivage. Les questions relatives aux droits d'auteur, à la propriété des données, aux lois sur la protection des données personnelles et aux politiques des plateformes influencent considérablement la constitution des collections issues des réseaux sociaux. Il est donc essentiel de documenter les processus décisionnels et de mettre en œuvre des politiques claires pour traiter et atténuer ces problèmes.