Ref. Ares(2026)8607867 - 10/09/2026

September 10, 2026

To: European Commission, Directorate-General for Education, Youth, Sport and Culture – Unit D1 for Cultural Policy

From: *ekip*, the European and Creative Industries Innovation Policy Platform (EU Grant agreement: 101112111)

Subject: New strategies to address impacts of AI scraping on gallery, library, archives and museum (GLAM) sector

ekip is an EU Horizon-founded research project and collaborative platform designed to boost the development of innovation capacity in European ecosystems to include the Cultural and Creative Industries (CCI) (see https://ekipengine.eu). We thank the European Commission for providing us with this opportunity to contribute evidence and recommendations on the implications of AI for the Cultural and Creative Sectors.

This Response provides information on how the gallery, library, archives and museum (GLAM) sector is engaging with AI, with a focus narrowed to the impacts on this sector by commercial AI scraping and specific challenges and recommendations around this. According to the most recent survey of data, GLAMs play a valuable economic role in the EU; there were 13.9 thousand GLAM enterprises in 2023 and 85.4 thousand people employed in the GLAM sector (European Union 2025). EU law recognises the priceless contributions of cultural heritage institutions to enhance access and participation in society and as arbiters for the transmission of European values (European Union 2022).

We therefore urge you to ensure that considerations for the GLAM sector are embedded in the strategy, both to strengthen the potential benefits of AI for cultural heritage and to ensure the retention of knowledge, creative potential, and underpinning infrastructures that makes the European cultural sector so vital both for those within the EU, and those engaging with it globally through digital experiences.

Submission prepared by:

Morgan Currie, Senior Lecturer; Science, Technology and Innovation Studies; ekip; University of Edinburgh

Lara Dal Molin, Research Assistant, ekip, University of Edinburgh

Caitlin McDonald, Research Assistant, ekip, University of Edinburgh

Nicola Osborne, Manager of the Institute for Design Informatics; ekip; and BRAID Creative Industries Lead, University of Edinburgh.

Matea Senkić, Research Assistant, ekip, University of Edinburgh

ekip response to call for evidence - Artificial intelligence strategy for the cultural and creative sectors

Anna Talley, Research Assistant, ekip, University of Edinburgh

Jennifer Volk, Research Assistant, ekip, University of Edinburgh

This report complements the response submitted by *ekip* to the European Commission's Call for Evidence on a better copyright environment (see: ekip team, The University of Edinburgh 2026)

High level overview: Threats of large-scale AI scraping to online cultural heritage

EU policy strongly recommends that knowledge, information and heritage be shared in an open way, using open access licenses wherever possible, while also recognising the importance of respecting rightsholders' interests (UNESCO, 2015; Council of Europe, 2005; European Parliament, 2019). To this end, over the past several decades, the European Union has supported a set of measures for digitising cultural heritage and making it widely accessible online (EU Commission, 2011; Cousins, 2018; Ziku and Fabos, 2022).

These infrastructures that sustain open digital heritage are now facing an acute and largely unacknowledged threat: since around 2022, and intensifying dramatically from 2024 onward, cultural heritage institutions have experienced a flood of automated bot traffic from artificial intelligence (AI) companies seeking to scrape data to train generative systems, such as Large Language Models (LLMs). This activity disregards the technical protocols that have governed the open web since its early years and inflicts real, measurable costs on public and taxpayer-funded institutions. The public heritage sector is, as a result, seeing a large transfer of value to private corporations largely without consent, attribution, or remuneration.

This response draws on 48 interviews and two zine-making workshops with 38 stakeholders from cultural heritage, GLAM, open culture and academic sectors as part of the University of Edinburgh's original and ongoing research. Current research and policy debates on AI and the cultural sector are largely dominated by copyright and licensing questions and risk overlooking the infrastructural and material dimensions of the crisis, and the specific vulnerability of publicly funded, digital heritage infrastructure. We spoke with 25 UK, US and EU institutions in-depth about how AI scraping is affecting their operations, finding shared experiences and concerns across geographic contexts. We also asked all who took part in the research about their ideas for a better path forward towards designing public interest, responsible access to valuable cultural collections, and giving the cultural heritage sector greater parity with the AI industry.

This response reports on our research findings in two parts: 1) **current harms of automated AI extraction on the cultural heritage sector** and 2) **this sector's ideas for future paths forward.**

In *Part 1, Costs to the cultural heritage sector of industrial-scale AI scraping,* we argue that the current model introduces financial, epistemic, and existential costs to the mission of public cultural heritage institutions that contribute open heritage material online. We identify three primary harms of industrial-scale AI scraping:

1. AI scraping in the manner and volume currently prevalent imposes substantial, unsustainable financial costs on cultural heritage institutions. Despite many cultural heritage institutions' willingness to provide their data in formats that are less costly on their infrastructure, they are still being scraped in ways that have costs on physical infrastructure, staff time, and institutional morale. Human traffic has also shown a decline to many of these institutions while their collection data is accessed via AI chatbot services, creating challenges for cultural heritage institutions in demonstrating their value.

2. The mitigation strategies institutions are adopting in response, particularly bot-blocking through cybersecurity services, are restricting public access while also closing valuable heritage data off from future AI models. This compromises the public mission many cultural heritage institutions serve, potentially reducing accessibility at a time when connecting with diverse audiences in inclusive ways is increasingly important.

3. Current modes of scraping ignore provenance and context, such that content may be presented by AI chatbots inaccurately. This poses a growing risk for misunderstandings – potentially contributing to disinformation, and/or misrepresentation – for minoritised groups and cultures and thus has the potential for inhibiting public trust in these institutions should their material be presented inaccurately by these third parties.

In *Part 2, Moving forward: dialogue between AI sector and cultural heritage/GLAM sectors,* we argue this stress on the infrastructures that sustain open cultural heritage presents an opportunity to reset the relationship between AI and the cultural heritage sector through technical, legal and governance frameworks that sustain openness while ensuring reciprocity and fair value return. We make six proposals for a way forward:

1. Compensation for infrastructure, not only content
2. New web standards for differentiated access
3. Protocols for provenance data to accompany AI extraction
4. Copyright laws that protect public-interest exceptions
5. Consensual, high-quality "scrapable" data environments
6. Investment in small AI and public-interest models

Part 1. Costs to the cultural heritage sector of industrial-scale AI scraping
Our research finds evidence that, while many in the cultural heritage and GLAM sectors see enormous potential for AI scraping of cultural data in terms of resource discovery and information retrieval, there are growing sectoral concerns. Our research finds three main implications of large-scale, consistent scraping for the sector that bear on this Response: 1) financial and infrastructural burdens on cultural institutions, 2) new online restrictions of heritage materials and open knowledge and 3) lack of context and control over information ingested by AI companies.

1. Financial and infrastructural burdens
GLAM institutions are facing monetary costs due to heavy bot traffic straining servers and impacting day-to-day operations. One institution told us, "*We are getting now in a day or in a week the amount of traffic we used to get in a year.*" In our research, this cost most often takes the form of additional labour hours required for web developers to mitigate slow or crashing servers. We also have evidence of institutions spending more on increased memory and disk space to improve service performance.

Institutions face an indirect cost through the decline in human traffic, as people who would have previously used these collections directly through the institution, now find information through AI chatbots and automated search box previews, causing challenges for institutions who are expected to demonstrate their value through evidence of engagement including online visitor numbers and usage data.

2. New online enclosures of digital cultural material
Paradoxically, heritage institutions' attempts to protect their service capabilities from bots and maintain open collections are creating new access restrictions around open heritage data. To mitigate costs, institutions are adopting software intermediaries – most commonly Cloudflare and Anubis – to prohibit bot activity from certain geographies or certain activities on their website, such as search, or disallowing bots altogether; this activity creates new firewalls that also block humans from some regions, therefore undermining aims to reach global audiences that many of these institutions hope to engage through their digitised collections.

Institutions' bot blocking further blocks their content from public interest usages of bots for research, art and web preservation, including for legal deposit libraries. Web archives we spoke to told us that web preservation has become much more complicated in recent months, as their crawlers are blocked by bot blocking software.

One critical knock-on effect of bot blocking is that high-quality, diverse, European heritage content is less accessible to AI companies' models, a loss which could have important implications for economic growth in the AI sector. For many people, LLMs are the new search engine; bot prevention blocks cultural heritage data from being incorporated into future AI models where many people now seek information.

3. Loss of context and provenance
AI bots ignore the carefully crafted APIs and bulk-access systems institutions build, instead scraping inefficiently via the user interface and ignoring standard data formats designed for ease of access and automated consumption. Curated context and structured metadata are therefore stripped away, such that content may be presented by AI models inaccurately. As one participant explained: “If you're a cultural heritage organization, the last thing you want to see is... everything that you put in your digital repository that's public domain winding up being sold back to people through scrapers and bots that have removed the context.”

Many GLAMs are currently taking actions to redress colonial ideologies and practices through their contextualising work, which is then lost when collection items become subsumed into LLMs that decontextualise these objects. Practitioners we spoke to expressed acute concern about this “decontextualisation” and its intersection with structural bias, such as bots absorbing photographs carrying a “colonial gaze” and reproducing these biases.

Part 2. Moving forward: calling for a dialogue between AI sector and cultural heritage/GLAM sectors
This emergent strain on public digital heritage infrastructure presents key opportunities to reposition the AI industry and cultural heritage sectors on more equitable and sustainable terms. Valuable public heritage content should continue to remain free, but those consuming infrastructure at industrial scale must be accountable and contribute to its sustainability. As one research participant noted, “A lot of the AI companies [have] realised that the organised, curated, provenanced, kept materials is really more valuable than they gave it credit for.”

Currently, attempts by institutions to enable dialogue or reach mutually beneficial agreements have largely been ignored: our interviews show that AI services providers have reached out only to the largest archives and institutions to forge explicit agreements for large-scale data provisions to train their models, leaving smaller institutions that are going to take the biggest hit from unexpected traffic spikes due to bot scraping outside the space of potential gains. Any policy response should account for differences among GLAM institutions, thus addressing issues of equity and participation.

We believe that the European Commission should demonstrate leadership in this space by supporting a range of policies for responsible and sustainable automated access to cultural heritage content. Below are six ways forward:

1. Compensation for infrastructure, not only content.
Drawing on the model of Wikimedia Enterprise (Becker 2026), policymakers should require large-scale commercial consumers of the commons to contribute to the infrastructure they burden through a levy on generative AI providers, mandatory contribution schemes, or cost-recovery mechanisms. This preserves open access to

content for the public while ensuring private beneficiaries pay for public infrastructure, directly addressing the “public cost, private benefit” asymmetry.

2. New web standards for differentiated access
We encourage support for and legal enforcement of emerging technical standards for specifying and enforcing types of access and consent, such as the IETF's AI Preferences vocabulary and Creative Commons Signals; these enable nuanced statements (e.g. “yes you can train on this document or collection, but with attribution/your rights expire and revert at a certain time”). Ideally, this work would interoperate with EU standards and laws to avoid fragmentation.

These standards should also embed FAIR and CARE principles in national digital collection standards. The CARE principles (collective benefit, authority to control, responsibility, ethics) provide a principled framework for the “as open as necessary” posture and for protecting fragile and community-held materials.

3. Protocols for provenance data to accompany AI extraction
We also call for provenance and preservation protocols for ensuring high-quality metadata accompanies extracted heritage material that appear in chatbot responses or AI generated search results. As one participant told us, “*I think we're in a position to provide more guidance and support and training in understanding research results that have been derived from AI*.” A stronger AI policy in this area for GLAMs can also ensure AI services retain important context around historic material that exhibit colonial perspectives, contributing to the 'protection and promotion of cultural and linguistic diversity and inclusion' as stated in the call.

Relatedly, LLM providers should also direct users back to source materials and provide institutions with usage analytics; this becomes especially important as institutions are also experiencing a loss of legitimate web traffic, both because services are rendered unavailable and because people increasingly obtain content from LLM services, rather than at the source. Such an approach would redress a context wherein these institutions are expected to justify their budgets through demonstrating public usage.

4. Reinforcement of copyright law to protect public-interest exceptions
The EU should modernise copyright and protect public-interest exceptions. Current copyright was not designed for machine engagement. New frameworks should provide enforceable protection against unauthorised scraping, while text and data mining (TDM) exceptions should remain calibrated to protect research, education and cultural heritage uses.

Material incorporated into models should be able to be removed upon request: this is of particular concern for archives holding personal and sensitive material. Whilst there are cases where this right is already protected through EU GDPR legislation, the lack of accountability and traceability in such models make this unfeasible. Although there are some protections under the EU AI Act, many of the types of AI models using GLAM and

archive materials may be perceived as low-risk systems, limiting the potential to regulate this use. Institutions therefore need a clear legal process when content is unlawfully scraped or infrastructure is damaged, and AI companies should maintain accessible contact points. Consideration should be given to a designated regulator or ombudsman function.

5. Support of consensual, high-quality "scrapable" data environments
Such environments would invest in curated datasets that bots are free to scrape, and their design should explicitly include commons and smaller actors, not only large rights holders. We note existing efforts in this area, such as the Common European Data Space for Cultural Heritage, which could facilitate mutually beneficial dialogue with the AI sector.

Institutions have a wealth of curatorial and provenance expertise from which AI development could greatly benefit, for appropriate remuneration and acknowledgement of their specialised, valuable skills. Where partnerships are pursued, policymakers should ensure institutional control and value return to public institutions. Any partnership models should be based around consensual training on content that institutions willingly provide, with fair compensation.

6. Investment in "small AI" and public-interest models.
Localised models could be developed for regional public benefit purposes, such as providing more materials in minoritised languages such as Catalan, Basque, Sami, etc. This approach offers a genuine alternative to the extractive paradigm, strengthens linguistic sovereignty, and supports the EU's ambitions to simultaneously foster cultural diversity and innovation.

Conclusion
The proposed constructive approaches will support the European Commission's goal for AI that sustains, rather than diminishes, access to culture and knowledge and contributes to more equitable relations between the AI and cultural heritage sectors in ways that will enhance both. This will require the AI sector to enter into dialogue with the GLAM sectors; AI that drives traffic back to sources; models enriched by cultural diversity; tools that improve metadata and accessibility; "small AI" and public-interest models; and, above all, cultural heritage institutions as active shapers of the technologies defining their collections' future.

We are happy to respond to any further questions about this document. Please contact Morgan Currie or Nicola Osborne at designinformatics@ed.ac.uk.

Works cited:

Becker, Lane (2026). The cost of 'free': How Wikimedia Enterprise protects Wikipedia in the AI era. July 16. Wikimedia Foundation. https://wikimediafoundation.org/news/2026/07/16/wikimedia-enterprise-protecting-wikipedia-ai/

Council of Europe. (2005). Convention on the Value of Cultural Heritage for Society (Faro Convention, 2005). https://www.coe.int/en/web/culture-and-heritage/faro-convention

Cousins, Jill (2018). Building a commons for digital cultural heritage. SMK – National Gallery of Denmark in Copenhagen (Statens Museum for Kunst). Available at: https://www.smk.dk/en/article/building-a-commons-for-digital-cultural-heritage/

ekip team, The University of Edinburgh (2026). Report on the review of the 2019 EU rules on copyright and related rights in the digital single market. Feedback reference F33504269. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/18172-Report-on-the-review-of-the-2019-EU-rules-on-copyright-and-related-rights-in-the-digital-single-market/F33504269_en

The European Parliament and the Council of the European Union (2019). Directive (EU) 2019/1024 of the European Parliament and the Council of 20 June 2019 on Open Data and the Re-Use of Public Sector Information, EU 2019/1024. https://eur-lex.europa.eu/eli/dir/2019/1024/oj/eng

European Union (2022). Council Resolution on the EU Work Plan for Culture 2023–2026 (2022/C 466/01). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32022G1207%2801%29&qid=1671635488811

European Union (2025). Culture statistics - cultural enterprises. December. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Culture_statistics_-_cultural_enterprises#Number_of_enterprises_and_value_added_by_product_category

European Union (2026). Government expenditure on recreation, culture and religion. August. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Government_expenditure_on_recreation,_culture_and_religion

UNESCO (2015). Recommendation concerning the Preservation of, and Access to, Documentary Heritage Including in Digital Form. UNESCO. https://unesdoc.unesco.org/ark:/48223/pf0000244675

Ziku M and Fabos B (2022). Digital Community Heritage and Open Access. CC Open Culture Working Group Digital Community Heritage, 30 November. Creative Commons. Available at: https://cc-digital-community-heritage.pubpub.org/pub/report

September 10, 2026.

Further reference

ekip (2026). Policy Areas: Digital Commons. Ekip website. https://ekipengine.eu/policies/digital-commons/

ekip (2026). *Creativity and civic infrastructure: CCIs and digital commons*. Scoping paper. Available at: https://ekipengine.eu/wp-content/uploads/2024/11/EKIP_digital-commons_Scoping-Note_final2-1.pdf