

"As Open as Possible, as Closed as Necessary": Balancing Community Needs, Sustainability, and Access in Transkribus

The mass-scraping of digital content by generative AI providers has precipitated a fundamental tension within the Digital Humanities. Many practitioners remain committed to the FAIR data principles — making code and data discoverable, accessible, interoperable, and reusable in pursuit of open science (Wilkinson et al. 2016; Vicente-Saez and Martinez-Fuentes 2018) — yet cultural heritage institutions and scholarly infrastructure providers are increasingly finding that unrestricted openness renders their platforms and collections vulnerable to predatory extraction (Weinberg 2025). Open license initiatives have, in some cases, inadvertently ceded collections wholesale to AI training pipelines (Cohen 2025), generating disquiet among open-access advocates and detractors alike (Eve 2025). The question this poster addresses is not whether openness remains a scholarly value — it does — but rather what form of openness remains defensible, and institutionally viable, under current conditions.

I argue that the principle of being "as open as possible, as closed as necessary" — a formulation originally developed within the FAIR framework to navigate GDPR constraints (Landi et al. 2020), and now increasingly adopted across the library and digital humanities sector (Bryant 2025) — offers a productive, if necessarily imperfect, framework for this recalibration. To substantiate this argument, this poster examines how READ-COOP, the European Cooperative Society that maintains and develops Transkribus (an AI platform for automated text recognition and information extraction from historical documents), has navigated precisely these opposing pressures since its establishment in 2019.

READ-COOP's business model requires selective closure of its infrastructure: the proprietary algorithms and extensive training datasets at the heart of Transkribus represent assets developed over years with substantial public investment, and without protection from re-appropriation by larger technology providers, the cooperative's viability — and with it, its capacity to serve its community — would be fundamentally threatened. This position has attracted criticism from researchers committed to open science approaches; and, indeed, it is important to acknowledge that such objections are not without force. Yet they tend to rest on principle rather than practicality. Few researchers possess the computational resources or technical expertise required to deploy the large models (approximately 400TB) that form Transkribus's core, were these to be openly licensed. The force of the objection is therefore largely symbolic — which does not make it irrelevant, but does affect how seriously it should constrain institutional decision-making.

At the same time, READ-COOP has pursued maximum accessibility through multiple mechanisms. Users retain full ownership of their data and may export all materials in open formats. Three hundred licensed AI models are available for internal community use, and members are actively encouraged to share training data through repositories such as HTR-United. Free monthly credits support hobbyist researchers; and the cooperative's Scholarship Programme has provided free access to students from 73 countries, supporting 336 individuals with over one million free credits by October 2024 (Nockels, Gooding, and Terras 2025). The cooperative structure itself embodies a form of openness that purely technical approaches

cannot replicate: democratic governance, monthly meetings, active communication channels, and annual conferences ensure genuine community participation in platform development and strategic decision-making (Terras et al. 2025). Transkribus may, in this sense, constitute a walled garden — but it is the Digital Humanities' walled garden.

READ-COOP's experience suggests that responsible platform development now necessitates pragmatic balance rather than an absolutist position on openness. This is not, of course, the only possible response to the current conjuncture: alternative models — from fully federated infrastructure to community-owned data commons — merit serious consideration. But the cooperative model, with its combination of selective closure, democratic governance, and broad accessibility mechanisms, offers one viable framework: one in which closure and openness function not as opposites but as strategically calibrated instruments for ensuring long-term institutional viability and community service. The broader implication, I suggest, is that the Digital Humanities community may need to reframe its expectations around openness — not abandoning the value, but rethinking what it can realistically mean in an age of unethical data extraction.

Bibliography

- R. Bryant (2025), "Open research as a strategic priority: Insights from an OCLC RLP leadership roundtable". *Hanging Together, the OCLC Research Blog*, 14th April 2025. <https://hangingtogether.org/open-research-as-a-strategic-priority-insights-from-an-oclc-rlp-leadership-roundtable/>
- D. Cohen (2025), "AI and Libraries, Archives, and Museums, Loosely Coupled". *DC Humane Ingenuity*, August 2025. <https://newsletter.dancohen.org/archive/ai-and-libraries-archives-and-museums-loosely-coupled/>
- M.P. Eve (2025), "Creative commons licenses and copyright may not stop academic work being used to train AI". *LSE Impact Blog*, 24th November 2025. <https://blogs.lse.ac.uk/impactofsocialsciences/2025/11/24/creative-commons-licenses-and-copyright-may-not-stop-academic-work-being-used-to-train-ai/>
- A. Landi, M. Thompson, V. Giannuzzi, F. Bonifazi, I. Labastida, L.O.B. da Silva Santos, and M. Roos (2020), "The 'A' of FAIR — as open as possible, as closed as necessary", *Data Intelligence* 2 (1–2), pp. 47–55.
- J. Nockels, P. Gooding, and M. Terras (2025), "Are Digital Humanities platforms facilitating sufficient diversity in research? A study of the Transkribus Scholarship Programme", *Digital Scholarship in the Humanities* 40 (Supplement 1), pp. 146–165.
- C.A. Romein, M. Terras, A. Stauder, B. Anzinger, and F. Stauder (2025), *Praeteritum transcriptum. A Transkribus Tribute: Celebrating our First Five Years as a Cooperative (2019–2024)*, READ-COOP SCE, Innsbruck. <https://zenodo.org/records/15308678>
- M. Terras, B. Anzinger, P. Gooding et al. (2025), "The artificial intelligence cooperative: READ-COOP, Transkribus, and the benefits of shared community infrastructure for automated text recognition" [version 2; peer review: 1 approved, 1 not approved], *Open Research Europe* 5:16. <https://doi.org/10.12688/openreseurope.18747.2>
- R. Vicente-Saez and C. Martinez-Fuentes (2018), "Open Science now: A systematic literature review for an integrated definition", *Journal of Business Research* 88, pp. 428–436.

M. Weinberg (2025), *Are AI Bots Knocking Cultural Heritage Offline?*, Engelberg Center on Innovation Law & Policy, NYU School of Law, v.1.0.
https://www.glamelab.org/files/Are_AI_Bots_Knocking_Cultural_Heritage_Offline.pdf

M.D. Wilkinson, M. Dumontier, I.J. Aalbersberg et al. (2016), "The FAIR Guiding Principles for scientific data management and stewardship", *Scientific Data* 3 (1), 15th March 2016.
<https://www.nature.com/articles/sdata201618>