

URGENT STATEMENT

Printed books, AI and rights: why we did not move forward with offers totaling hundreds of thousands of dollars, and what we propose to the publishing sector



Several offers from major technology platforms, mediated by confidentiality agreements and jointly amounting to significant sums, have ceased to be mere commercial transactions. In the artificial intelligence economy, they are also the gateway to a much deeper discussion: digital reproduction, text and data mining, content traceability, and the actual capacity of rights holders to establish conditions of use over their works. That is why, at Hipertexto - Netizen, we decided not to move forward — but we did decide to act.

We are making this decision public because the issue is greater than any single company: it concerns the entire publishing sector. What is at stake is the sector's position with respect to operations that can no longer be interpreted through the frameworks of just a couple of years ago. The printed book has also entered the AI conversation, and the book ecosystem needs to respond with greater clarity, better signals, and an urgent agenda for protection and coordination.

WHAT WE BELIEVE AND PROPOSE FROM HIPERTEXTO - NETIZEN

Halting offers from technology platforms and high-risk sales because they can no longer be assessed as simple purchases of printed copies, but also for their implications regarding content. Buying a printed book does not amount to a license to reproduce it, scan it, or incorporate it into

text and data mining or model training processes.

This is not an isolated case: a substantial part of the publishing sector continues to enter these conversations without visible, verifiable, and machine-readable use reservations. From Hipertexto - Netizen (HN), we build tools, standards, and signals leveraged through SIMEH (Integrated Standardized Metadata System) so that publishers can declare restrictions, strengthen traceability, and make their conditions of use visible to both people and automated systems.

We call upon the sector and trade associations across Ibero-America to move from scattered concern to a coordinated response: review catalogues, reinforce metadata, activate reservations regarding the use of works in text and data mining and AI training (TDM), update rights pages, and generate actions aligned with current needs.

Why we halted these sales

At Hipertexto - Netizen, we reviewed offers from major technology platforms to acquire printed books. Although they represented a highly significant commercial opportunity, we decided not to move forward because they raised a fundamental question: what could happen to those copies once they left the publishing circuit and entered a logic of mass acquisition?

From a legal standpoint, this was no longer a simple sale. When a distributor sells a book to a reader, a library, or another bookstore, the transaction is clear and limited: ownership of the physical copy is transferred and the exhaustion of the distribution right applies. However, when there is reasonable suspicion that the purpose may be mass digital reproduction to train models, the operation changes in nature and edges into a far more uncomfortable scenario: the conscious facilitation of a use that could compromise reproduction rights that do not belong to the distributor, but to publishers and authors.

Added to this is a good-faith concern. In the analog world, we distribute printed books, but we are not the holders of their reproduction rights. Handing over copies knowing — or treating as irrelevant — that they could be instrumentalized in an unauthorized reproduction operation would be, at the very least, ethically questionable and legally risky vis-à-vis the publishers who have entrusted their catalogues to Hipertexto - Netizen.

Cases such as *Bartz v. Anthropic* — currently one of the largest ongoing lawsuits over the mass use of books to train models — illustrate the scope of harm these operations may cause to authors, publishers, and the sector as a whole. Contributing to such a chain, even indirectly, was incompatible with what we have been publicly defending.

The printed book has already entered the AI discussion

The debate on AI and publishing rights tends to focus on digital files, web repositories, and online catalogues. That reading has fallen short. The printed book is also at stake: it can be acquired, digitized, transformed, and fed into large-scale processing flows. The materiality of the copy does not eliminate the possibility of subsequent reproduction of its content.

The distinction is not minor. Acquiring a physical object is not equivalent to acquiring the right to reproduce the work it contains. Confusing the circulation of copies with authorization over the content is a dangerous oversimplification, and in the current environment it may end up weakening the position of authors, publishers, and rights holders.

The analog and the digital are no longer separated by a clear boundary. Protection of the catalogue cannot begin when the file enters the web: it must begin earlier — on the rights page, in the metadata, in public use policies, in reservation records, and in the signals that accompany the work regardless of its format.

What is unprotected today in the sector

The sector's main deficit is not one of awareness; it is one of infrastructure. Many publishers know the problem exists, but they still lack consistent mechanisms to express, document, and make discoverable their conditions of use. In too many cases, the rights holder's will depends on scattered texts, tacit decisions, or clauses that do not communicate with the automated systems that today crawl, aggregate, and process content.

There is also a lack of common language. The publishing ecosystem continues to speak with too much ambiguity about digital reproduction, TDM reservations, traceability, licenses, content identification, and proof of the rights holder's will. That ambiguity makes defense more expensive and extraction cheaper: as long as there is no clear and shared vocabulary, negotiation will remain asymmetrical.

To this is added a regional weakness. No publisher will solve alone a problem that is already transnational. Technological pressure does not stop at a border, a book fair, or a catalogue. Without shared responses among publishers, distributors, associations, platforms, and book-sector organizations, the industry will continue to react late and in a fragmented manner.

What we are doing at Hipertexto - Netizen

At Hipertexto - Netizen, we have moved from warning to action. We have taken an active sectoral position in building infrastructure so that publishers can declare restrictions, strengthen traceability, and make their conditions of use visible to both people and automated systems. This is the logic that underpins SIMEH, Book2Data — the SIMEH module that automatically extracts ONIX metadata from PDF, EPUB, or Word files — and the technical and legal layers we have been developing around metadata, TDM reservations, and proof of the rights holder's will.

In practice, this work includes the implementation of restrictions in ONIX (the international standard for publishing metadata that accompanies the book throughout the distribution chain), the incorporation of machine-readable TDM signals, the use of content-derived identifiers such as the ISCC (International Standard Content Code, ISO 24138:2024) — an international standard that generates a digital fingerprint from the file and enables verification of the work regardless of its name, format, or hosting; the enrichment of ingestion and cataloguing flows through Book2Data; and the construction of publicly accessible registries. It also includes the web layer: pages, headers, visible policies, and monitoring and control tools so that the publisher's declaration does not depend on an isolated text, but rather on a coherent system of signals.

That effort is already publicly available at SIGNAL, <https://signal.simeh.co>, the public space managed by publishers within SIMEH regardless of license type, where any person or automated system may consult declarations and TDM conditions of use for registered works, in both digital and print formats, with certified date and time. SIGNAL is not a promise: it is up and running, open, in continuous development, and available to the publishing sector.

Our aim is to help the sector move from a defensive discourse to a genuine technical capacity to document, publish, and make discoverable its reservations. If publishing content has value in the AI economy, that value requires infrastructure commensurate with the problem.

Tools requiring immediate attention

- ONIX metadata with clearly declared conditions of use.
- Machine-readable TDM reservation signals and protocols.
- Content-derived identifiers, such as ISCC, to strengthen persistence and traceability.
- Publicly accessible registries, such as SIGNAL, for declarations and conditions of use.
- Visible policies in catalogues, publisher pages, headers, and rights pages.

What we urgently propose to the publishing sector

The first step is to stop treating this discussion as if it were exclusively legal or exclusively technological. It is both at once. That is why we propose a sectoral response that combines rights, metadata, traceability, public policy, and institutional coordination.

The immediate steps are concrete: publishers must review their catalogues, strengthen the way they declare conditions of use over their works, update metadata, activate machine-readable reservations, review rights pages, and reinforce registries or equivalent mechanisms that allow verifiable documentation of the rights holder's will. Every month that passes without taking this step widens the sector's vulnerability.

But this cannot remain confined to isolated publishers. We call for a much broader and better-coordinated industry conversation, with the active participation of:

- Ibero-American book chambers and their national counterparts.
- CERLALC and multilateral organizations focused on books and reading.
- Associations of independent and university publishers, and the Ibero-American Publishers Group.
- Distributors, bookstores, book fairs, and commercialization platforms.
- National libraries, university networks, and documentation centers.
- Author associations, collective management societies, and rights holders.
- Law firms and legal advisors specialized in intellectual property and AI.

The problem is already too large to continue addressing it in silos. We need a public implementation agenda with minimum common recommendations, regional criteria, and an adoption roadmap so that the sector arrives better prepared for future negotiations with major technology platforms and AI providers.

A cultural decision is also required: to stop arriving late. Reacting once the operation is already on the table is not enough. We must arrive earlier, with clear criteria, visible reservations, and explicit conditions of use.

The difference between negotiating from a position of clarity or from technical and documentary emptiness is already, today, a strategic difference for the sector.

SECTORAL WEBINAR · WEDNESDAY, JULY 15, 10:00 A.M. (BOGOTÁ TIME)

We invite the Ibero-American publishing sector and book industry stakeholders to a webinar on **Wednesday, July 15, at 10:00 a.m. (Bogotá time)**, with the participation of specialized legal counsel, to explain this initiative in detail, present **SIGNAL**, address questions, and propose shared roadmaps that may be endorsed, supported, and deployed by trade associations.

Time equivalencies: Mexico City 09:00 a.m. · Bogotá / Lima / Quito / Panama 10:00 a.m. · Santiago / Asunción / Caracas / La Paz / San Juan 11:00 a.m. · Buenos Aires / Montevideo / São Paulo 12:00 p.m. · Madrid / Lisbon 05:00 p.m.

Registration open: <https://shorturl.at/Dl9Zj>

Artificial intelligence is here; the publishing sector cannot respond with either denial or resignation. What it can do is arrive on time to the discussion with better signals, better negotiating conditions, and a clearer understanding of what it is defending.

If publishing content has value in the AI economy, that value can no longer be managed from opacity, ambiguity, or the assumption that silence equals consent.

The time to act is now. With greater clarity. With greater traceability. With a coordinated regional response. And with the conviction that defending the sector's rights is not a gesture against innovation: it is the condition for innovation to be legitimate.

References

RECENT PUBLICATIONS BY HIPERTEXTO - NETIZEN

- Those who remain silent, train: Hipertexto - Netizen, publishers and associations in dialogue with AI — hipertexto.com.co
- Hipertexto - Netizen AI Declaration: responsible use in the publishing ecosystem — hipertexto.com.co
- Publishing sovereignty in the era of AI — hipertexto.com.co
- SIGNAL - Public registry of TDM declarations for publishing catalogues — signal.simeh.co

ANALYSIS AND OPINION

- The destruction of the analog book, the ruin of culture — msn.com
- Buying books that have been sold and digitizing them to train AI — linkedin.com
- Arantza Mellado, Will ONIX continue to be sufficient for discovering books in the age of AI? — actualidadeditorial.com

INTERNATIONAL PRECEDENTS

- Bartz v. Anthropic: the largest copyright settlement in the U.S. concerning the use of books in AI — legalblogs.wolterskluwer.com
- Anthropic wins on fair use for training its LLMs; loses on building a central library of pirated books — authorsalliance.org

INITIATIVES AND TECHNICAL STANDARDS

- Bookwire and Liccium: innovation to protect content from AI use — bookwire.net
- Liccium: TDM opt-out declarations based on ISCC (ISO 24138:2024) — docs.liccium.com
- ISCC — International Standard Content Code (ISO 24138:2024) — iscc.codes