

Seven Questions about CC Signals

By **RICK ANDERSON** MAY 26, 2026

ARTIFICIAL INTELLIGENCE | AUTHORS | BUSINESS MODELS | COMMERCE | CONTROVERSIAL TOPICS
| COPYRIGHT | DATA PUBLISHING | ECONOMICS | ETHICS | OPEN ACCESS | OPEN SCIENCE
| SOCIAL ROLE | TECHNOLOGY | TOOLS | USABILITY | WORLD OF TOMORROW

As most *Scholarly Kitchen* readers will know, in 2002 the [Creative Commons Organization](#) (CCO) created a suite of licenses that copyright holders can apply to their works in order to make them available for free reuse by the public. Different licenses grant permission for everything from functionally unlimited reuse ([CC BY](#)) to a much narrower range of uses ([CC-BY-NC-ND](#) or [CC-BY-NC-SA](#)), and once applied the licenses are [pretty much irrevocable](#).

CC BY creates a [blanket permission](#) for the work to be “translated, altered, arranged, transformed, or otherwise modified in a manner requiring permission under the Copyright and Similar Rights held by the Licensor” by anyone, in any manner and for any purpose, as long as the creator of the original work is acknowledged. It’s generally considered the only license that makes a work fully open access (OA), since the other licenses impose restrictions that go against key principles outlined in the [Budapest Open Access Initiative statement](#). For this reason, ensuring that scholarly products such as articles, books, and data sets will be made available on a CC BY basis (or the equivalent) has been a broadly shared goal within the OA movement for decades.



The Tyranny of Unintended Consequences

There are three certainties in life: death, taxes, and unintended consequences. One unintended consequence of 25 years of pervasive and international pressure to make scholarly publications and data sets freely available for functionally unlimited reuse is that millions of scholarly products are now being used enthusiastically by (among others) a growing horde of artificial intelligence (AI) agents in ways that make the copyright holders in those products uncomfortable. Those who have applied irrevocable CC BY licenses to their works, however, have no obvious legal recourse.

It's worth pointing out, though, that although the availability of these scholarly products for exploitation by massive commercial AI companies may have been *unintended* on the part of the works' creators, this outcome was not *unforeseen*. On the contrary, the potential for machine learning to take advantage of open content has been on our collective radar for decades, and was in fact explicitly celebrated by OA advocates from the beginning of the movement. As early as 2006, John Willinsky [was arguing for](#) the "automated indexing of the scholarly literature" and for "automated systems (for) citation rankings of article and journal" – both of which would necessarily be AI functions based on large-scale ingestion of text – and saw these as part and parcel of a shift to open access. As recently as five years ago SPARC was [happily anticipating](#) the use of "a modern machine learning, natural language processing approach" to analyze how federal data sets are used across "millions" of publications and thereby "demonstrate the value of data as a strategic asset." SPARC's website continues to [champion](#) the development of "powerful text and data mining tools that can analyze the entire research literature, uncovering trends and connections that no human reader could," and [advocates](#) for open data policies that "(permit) any user to download, copy, analyze, re-process, pass to software, or use (the data) for any other purpose." The benefits of such openness are both clear and considerable; some of the costs and downsides, however, have also been [clear to many for some time](#), but are now apparently becoming clearer to those who have up until now been unreservedly advocating for it. (See, for example, the very different consideration of these implications in SPARC's 2024 news item titled "[Reclaiming Control: Privacy, Platforms, AI & Governance in the Public Interest](#)", in which they call on "institutions [to] negotiate more assertively to limit data collection and non-academic use of academic data.")

Creative Commons Starts Working on a Solution

In June 2025, CCO took a step in the direction of helping authors try to rein in the massive reuse of their CC-licensed products by [introducing CC Signals](#), a "new preference signals framework designed to increase reciprocity and sustain a creative commons in the age of AI." Framed as "a major step forward in building a more equitable, sustainable AI ecosystem rooted in shared benefits," CC Signals is intended to give copyright holders a mechanism for signaling how they would (and would not) prefer that their CC-licensed content be reused.

Without getting into [the legal technicalities](#), the fundamental limitation of this approach is obvious: once they've made their work CC BY, copyright holders' preferences no longer carry any weight. The right to reuse the licensed content in functionally unlimited ways has been irrevocably granted to the general public, including those who operate AI agents and LLMs, and the license requires nothing of the reuser except acknowledgment of the licensor as original creator — and even then, only when a genuinely derivative work has been created. And once the license has been granted the licensor can't change his mind and restrict the general public's rights in the work, either subsequently or retroactively. Despite some of the language contained in the current version of the CC Signals framework (referring to it, for example, as "a set of criteria that AI developers must meet"), the signals ultimately constitute requests, not legally-binding criteria that anyone "must meet."

As part of its ongoing effort to address this issue – and after soliciting public input and feedback on the initial CC Signals framework – last week the CCO announced an [update to the program](#). Framing it as a suite of "several high-impact interventions... to restore trust, strengthen participation, and embed public interest values into the AI knowledge ecosystem," this new effort will "define best practices for attribution in AI contexts," propose and advocate for "the development and usage of carefully scoped AI opt-outs," and seek to develop "a new tool designed to enable conditional access to openly shared collections and compilations."

Seven Questions for the Creative Commons Organization

These efforts raise many, many questions, of course, as CCO has acknowledged; the organization is clearly working very hard to address the most important of these.

I reached out to CCO multiple times, asking if someone would be willing to respond to a few questions about the new CC Signals framework, and received no response. So since an interview isn't possible, I'm going to pose the questions here and solicit input from our readers. How do you think these issues will or should be addressed?

1. **Public interest use.** The distinction between "public interest use" and other kinds of use is repeatedly invoked in the CCO's discussion of issues related to AI-based use of open content. But nowhere is "public interest" defined. What distinguishes a "public interest use" from any other use of a CC-licensed work? In this context, does the phrase simply refer to not-for-profit use, or will the new CC Signals framework

contemplate narrower criteria? (And of course, it bears pointing out that nowhere in CC license language is there any requirement that reuse of the licensed work be undertaken in the public interest.)

2. **Honoring the agency of creators (1).** CCO is advocating for “the development and usage of carefully scoped AI opt-outs that simultaneously sustain creator agency while protecting public interest uses.” But hasn’t the creator exercised agency, in the first instance, by choosing to grant the public an irrevocable license to make free and unrestricted use of the work? After that license has been granted, on what basis should the creator expect to exercise agency over others’ use of the work? (Perhaps CCO is planning to create a new version of the CC BY license that incorporates such opt-out language — but for the millions of articles, books, and [copyrightable data sets](#) already licensed under its existing terms, the horse has left the barn; there is no “opting out” for those works, and rights granted by existing CC licenses can’t be narrowed retroactively.)
3. **Honoring the agency of creators (2).** CCO notes that feedback it received “was direct and consistent in stating that preference signals without enforcement do not meaningfully shift power. Signals alone cannot create agency in a system that many people did not choose to participate in.” But again, there’s no need to “create agency” in this system — under most copyright regimes, the moment a work is created agency is fully and automatically vested in its creator, who then chooses for herself whether or not to license her work (unless, of course, she’s been [compelled to do so](#) by a third party). “Power” over the work “shifts” precisely when the copyright holder exercises her agency by relinquishing control over her work and handing it over to the general public — and CC licenses are explicitly designed to prevent any shift in power from the public back to the copyright holder.
4. **Attribution.** Given that CC BY licenses require acknowledgment of the original creator when the work is reused or derivatives are created, what would be a reasonable threshold for acknowledgment when an AI agent creates a document that contains duplication or derivatives of by CC BY licensed content? For example, if a bot formulates a query response that includes duplicated or derivative content from a 10,000-book corpus that includes one CC BY book, does the license require the resulting output to acknowledge the author of that book? On the other hand, what if the bot quoted or created derivatives from 10,000 books, 5,000 of which are CC BY, each with a different author? Would the resulting output need to provide attribution for all 5,000 authors?
5. **“Responsible” reuse.** CC notes that “many want to continue sharing their collections while ensuring that AI developers use them responsibly by respecting attribution, ensuring transparency, and meeting other safeguards aligned with their public interest missions. We want to build tooling to enable this in standardized, legally enforceable ways.” Here the question is similar to that posed above regarding “public interest” use. Who will define “responsible,” and how will CC Signals account for differing views as to what constitutes “responsible” reuse of openly licensed content? Legal technicalities aside, is it even within the spirit of openness for a work’s creator to define “responsible use” in a way that is binding on others?
6. **Applying licenses to non-copyrightable products.** CCO’s guidelines on “[Using CC-licensed Works for AI Training](#)” indicate that “when training data is subject to the [ShareAlike condition](#), model outputs and the model itself, if shared publicly, should be made available under the same CC license as the original works.” But how can AI developers or users apply any license at all to outputs created by AI agents, given that only the copyright holder can apply a CC license and works created by AI agents are [not copyrightable](#) (in the U.S., anyway)? When a work has no copyright holder, no one has authority to apply a license to it or otherwise restrict its reuse.
7. **Community control of data.** CCO’s update indicates that the organization’s “north star remains the same: sustain access to human knowledge. Today, that means more than enabling sharing. It means questioning long-held assumptions, and ensuring communities are in control of their own data.” What assumptions is CCO referring to? And in particular, in what way is open access to knowledge sustained when a group can exert control over “(its) data”? It may be the case that all things considered, certain groups (which ones?)

should retain control over data related to them – but surely such control would represent a compromise in terms of open access to human knowledge, not a manifestation of openness.

To our readers: if you were the Creative Commons Organization, how might you answer these questions? And what other questions should we be asking?



Rick Anderson

Rick Anderson is University Librarian at Brigham Young University. He has worked previously as a bibliographer for YBP, Inc., as Head Acquisitions Librarian for the University of North Carolina, Greensboro, as Director of Resource Acquisition at the University of Nevada, Reno, and as Associate Dean for Collections & Scholarly Communication at the University of Utah.