S A Y A R I

Via Electronic Submission

30 October 2026

European Data Protection Board
Rue Wiertz 60
B-1047 Brussels
Belgium

Re: Public Consultation on Guidelines 03/2026 on Web Scraping in the Context of Generative AI — Publicly Accessible Data Published for Third-Party Reliance

Dear Members of the Board:

I. Introduction

Sayari Labs, Inc. ("Sayari") submits these comments on Guidelines 03/2026 on web scraping in the context of generative AI, adopted 7 July 2026 (the "Guidelines").[1] We confine ourselves to a single point.

Sayari supports the Board's decision to address web scraping directly, and we accept without reservation the premise that public accessibility is not itself a lawful basis. Our concern is with the starting point the draft adopts in its place. Section 3.2.3 builds the reasonable-expectations analysis almost entirely out of materials drawn from social media and content-sharing platforms, and then applies the result to the whole of the open web. For the category of data that matters most to the Union's own regulatory objectives — corporate and beneficial ownership registers, official gazettes, court and insolvency records, regulatory filings, procurement awards, customs and trade records, and sanctions and enforcement designations — that starting point is not merely imprecise. It is inverted. The draft assigns the weakest expectation of further processing to precisely the data that is published so that third parties will process it.

Publicly accessible data published for third-party reliance requires a different starting point. Sayari asks the Board to adopt a rebuttable presumption:

> Where data has been made publicly accessible in order that third parties may inform themselves of the matters published, and the controller's purpose is compatible with that purpose, the processing falls within the data subject's reasonable expectations unless the particular circumstances indicate otherwise.

This is not a request for an exemption. The presumption is rebuttable; the Article 6(1)(f) balancing test continues to apply in full; Article 21 remains intact. What it changes is which party bears the burden, and on what the analysis turns. Sections II and III explain why the draft's present starting point cannot bear the weight placed on it, and Section IV sets out the amendment we propose.

II. Sayari's Interest

Sayari is the judgment infrastructure for trustworthy AI in economic security and commercial risk. The Sayari Commercial World Model resolves 10.6 billion primary-source records from 250+ jurisdictions into a connected graph of companies, people and trade activity. Our customers are government agencies,

financial institutions and corporates who use that graph to identify beneficial owners, trace ownership and control for sanctions purposes, and map supply chains for forced-labour and export-control compliance. A Judgment Ontology, encoding over 11 years of investigative tradecraft, and Superconductor, an agentic orchestration platform, sit on top of it.

III. The Draft Mistakes the Preferences of Website Operators for the Expectations of Data Subjects

The third factor in the draft's balancing analysis asks whether the data subject has taken steps to restrict access, and treats robots.txt directives, CAPTCHA, login walls and user-agent restrictions as relevant evidence of what the data subject can reasonably expect.[4] The difficulty is that in this domain the data subject did not place the data online, does not control the server, and has no relationship with whoever configured it.

No company director has read the robots.txt file of the register in which her appointment is recorded. No person named in an insolvency notice chose the gazette's crawl-delay. Those files express the interests of the entity operating the site — bandwidth management, commercial licensing of the same data to paying subscribers, and in some cases the exclusion of competitors. Those interests are frequently adverse to the people in the database. A register operator that sells bulk extracts has a commercial reason to block crawlers; that reason has nothing to do with the privacy of the directors it lists, and it should not be allowed to masquerade as their expectation.

The draft already concedes half of this. Paragraph 45 correctly states that the absence of technical restrictions cannot be read as consent to scraping.[5] The corollary follows necessarily: if the absence of a signal from the operator tells us nothing about the data subject, the presence of one cannot tell us anything either. The same actor is sending the same kind of message in both cases. As drafted, the factor operates as a one-way ratchet in which the operator's silence is neutral but the operator's objection is attributed to a third party who never made it.

The deeper error is one of axis. The draft locates reasonable expectations in the volition of the data subject — did this person choose to put this information where it could be found. That is the right question for a social media profile. It is unanswerable for a register entry, because the data subject had no choice: the disclosure was compelled by law, often as a condition of limited liability. Reading volition as the measure of expectation means that the more firmly the legislature has decided the public needs to know something, the less the publisher may expect it to be used. The Court has not taken that view of publicly available data, and the Board should not introduce it here.[6]

The correct axis is purpose. Where the law requires publication so that counterparties, creditors, journalists and regulated firms may inform themselves, a person recorded in that publication can reasonably expect that those parties will in fact inform themselves — including systematically, and including by machine. What the data subject may reasonably expect is set by the purpose for which the data was made public, not by the technical access conditions imposed by whoever happens to host it.

IV. The Amendment We Propose

Sayari recommends two amendments. The first is the substance of this submission; the second conforms the third factor to it.

A. A new paragraph following paragraph 64. The Board should add:

> Where personal data has been made publicly accessible by a public authority, by a person acting under a legal or regulatory obligation of disclosure, or otherwise in order that third parties may inform themselves of the matters published — including corporate and beneficial ownership registers, official gazettes, court, insolvency and enforcement records, regulatory and listing filings, procurement awards, customs and trade records, and sanctions designations — the data subject's own volition is not the measure of what the data subject may reasonably expect. The measure is the purpose for which the data was made public. Where the controller's purpose is compatible with that purpose, the processing should ordinarily be regarded as falling within the reasonable expectations of the data subject. This presumption may be displaced where the data concerns the data subject in a private rather than a commercial, professional or official capacity, where the processing is directed at the individual rather than at the matters disclosed, or where other circumstances specific to the data subject so indicate.

B. Amendment to the third factor at paragraph 64. After the reference to technical restrictions, the Board should add:

> Restrictions imposed by the operator of a source — including robots.txt and ai.txt directives, CAPTCHA, user-agent discrimination and registration requirements — are relevant to the interests of that operator. Where the data subject is not the operator and did not place the data on the source, such restrictions are not of themselves evidence of what the data subject can reasonably expect, and should be weighted accordingly.

Two consequential points follow, and we flag them briefly. Paragraph 66(c), which offers the exclusion of registers and official sources from collection as a mitigating measure, should be clarified so that it does not apply where the controller's purpose is the verification, identification or integrity checking of the entities to which the data relates; as written it contradicts paragraph 42.[7] And paragraph 66(f), which suggests a prior opt-out facility, is self-defeating in this domain: a published list of persons who wish not to appear in sanctions, insolvency or beneficial ownership screening would itself be an intelligence product of considerable value to those persons' counterparties. Article 21(1), assessed case by case, is the mechanism the Regulation provides, and it is sufficient.[8]

V. On the Limits of the Position We Advance

We anticipate that the Board will weigh this submission against its judgment in matters such as Sovim, in which the Court invalidated the provision of the Anti-Money Laundering Directive granting the general public access to beneficial ownership registers, holding it a disproportionate interference with Articles 7 and 8 of the Charter.[9] We do not read Sovim as an obstacle to the amendment we propose, and the distinction matters.

Sovim concerned the conditions under which a Member State must make a register open to all comers, without any statement of purpose. It did not hold that beneficial ownership data is beyond lawful processing; the Court expressly preserved access for those able to demonstrate a legitimate interest, naming obliged entities, journalists and civil society organisations working on money laundering and terrorist financing. Our proposal is framed in exactly those terms. It is available only where the controller's purpose is compatible with the purpose of publication, it is rebuttable on the data subject's particular circumstances, and it leaves Article 21 intact. What it removes is not the data subject's protection but the draft's reliance on a website operator's crawl directives as a proxy for it.

Nor do we ask the Board to depart from Satakunnan Markkinapörssi, which holds that data lawfully in the public domain under a national publicity regime remains within the scope of data protection law while accepting that the purpose of republication bears materially on the assessment.[10] That is our position. Manni supplies the other half of it: there is no general right to erasure from a companies register, because the legal certainty of third parties dealing with companies justifies continued publicity, and that justification runs to the people who rely on the register as much as to the register itself.[11] A register nobody is permitted to read at scale does not deliver the legal certainty the Court described.

VI. Conclusion

The difference between a social media post and a companies register is not a matter of degree. One is a person telling the world about themselves. The other is the state telling the world about a company, because the law concluded that the world needed to know.

Sayari would welcome the opportunity to discuss these points with the Secretariat or the relevant expert subgroup, and will provide whatever operational detail on registry-based diligence would assist the Board's work.

Respectfully submitted,



Owen Denby, General Counsel & Secretary
Sayari Labs, Inc.
1152 15th Street NW, Suite 910
Washington, DC 20005
Tel: +1 (202) 621-9821
Legal@sayari.com

Notes

1. EDPB, Guidelines 03/2026 on Web Scraping in the Context of Generative AI, Version 1.0 (adopted 7 July 2026), https://www.edpb.europa.eu/public-consultations/guidelines-032026-on-web-scraping-in-the-context-of-generative-ai_en.

2. Guidelines, ¶¶ 42, 66(c).

3. Id. ¶¶ 1, 4.

4. Id. ¶ 64, third factor.

5. Id. ¶ 45.

6. See Case C-252/21, Meta Platforms Inc. and Others v. Bundeskartellamt, ECLI:EU:C:2023:537 (4 July 2023), on which the Guidelines rely, and in which the Court held that data is "manifestly made public" under Article 9(2)(e) only where the data subject has by an explicit and intentional act made it accessible — a volitional requirement appropriate to that derogation and inapposite to Recital 47.

7. Guidelines, ¶ 66(c), second indent; cf. id. ¶ 42.

8. Id. ¶ 66(f).

9. Joined Cases C-37/20 and C-601/20, WM and Sovim SA v. Luxembourg Business Registers, ECLI:EU:C:2022:912 (22 Nov. 2022).

10. Case C-73/07, Tietosuojavaltuutettu v. Satakunnan Markkinapörssi Oy and Satamedia Oy, ECLI:EU:C:2008:727 (16 Dec. 2008).

11. Case C-398/15, Camera di Commercio, Industria, Artigianato e Agricoltura di Lecce v. Manni, ECLI:EU:C:2017:197 (9 Mar. 2017).