

# Response to the Conduct Market Authority's consultation on proposed conduct requirements for Google's general search services

25 February 2025

*Submitted by the Open Data Institute, as part of the CoCoDa project*

## Our approach to this consultation response

We recognise the CMA's designation of Google with 'strategic market status' and the drafting of requirements as a step in the right direction. However, these requirements do not go far enough.

We consider the top priorities for the CMA on Google's dominance Online Search as:

1. Responding to Google's continued moves towards market dominance in Search and AI
2. Introducing meaningful data access remedies beyond data portability
3. Strengthening attribution transparency in generative AI
4. Guaranteeing real-time, interoperable data portability
5. Upgrading compliance monitoring and reporting

### ODI

This response is led by the [Open Data Institute](#) (ODI), a non-profit organisation based in London, as part of a wider project called CoCoDa, a collaboration between the University of St Gallen, the University of Lausanne, Maastricht University and the ODI.

It is built on several years of experience at the ODI examining data portability mechanisms, including the stewardship of [Solid](#) with the help of funding from Google.org, with Solid being an open data standard focused on giving people control over their personal data. The ODI have also been involved with numerous initiatives related to Smart Data in the UK, for example stewarding the Smart Data Challenge, and involvement in Open Banking initiatives. Further, via the CoCoDa project we build on interdisciplinary examinations of systemic risks that emerge from the dominance of online platforms and their accumulation of data - with an emphasis on legal mechanisms and the technical infrastructure that is needed for both examining and redressing these systemic risks, in particular Article 40 of the EU's Digital Services Act (DSA).

### CoCoDa

[CoCoDa](#) (Concentration of Control and Data) is an interdisciplinary research project funded by the Swiss National Science Foundation under the International Co-Investigator Scheme, which brings together technical and legal expertise from the University of Lausanne, the University of St Gallen, Maastricht University (NL), and the Open Data Institute (UK).<sup>1</sup> Our research examines how the concentration of control and data in dominant online platforms

---

<sup>1</sup> See: <https://snsf-cocoda.github.io/>.

creates systemic risks in society, and we develop techno-legal solutions that make legal transparency mechanisms operational and enforceable in practice.<sup>2</sup>

## Response

Assessing the impact and demands associated with Google's dominance is of great importance in the present moment. The dominance of Google services extends beyond their own services; the company has completed deals such as a recent agreement for the integration of Gemini models into new iterations of Apple's Siri<sup>3</sup>. Further increases in the concentration of online infrastructure controlled by Google<sup>4</sup> mean that there is an even greater need to address UK citizen's access to information and publishers' ability to provide factual content.

## Publisher Conduct

### **6.4(b) We welcome further examples of information and metrics that help explain how Search Content is attributed in search generative AI features and the factuality of those features, and views on how these data would be best disseminated**

We welcome the CMA's requirement that Google publish its approach to attribution and factuality. Further to what has been proposed within the Publisher Conduct CR, we believe that this could be strengthened in a number of ways.

As the most important element, we consider the pursuit of effective data remedies for researchers and competitors, as already existing under the DMA and DSA in the EU. One particularly important focus should be click-and-query data, which already must be provided in the EU but requires further investigation to ensure this mechanism is effective. Click-and-query data is instrumental for the development of AI tools. This is demonstrated by the suggestion that ChatGPT's Retrieval-Augmented Generation (RAG) has seemingly used and may continue to use Google Search when providing answers on real-time topics.<sup>5</sup>

---

<sup>2</sup> See: Bekavac, L., Strecker-Bischoff, J., Garcia, K., Mayer, S., & Tamò-Larrioux, A. (2026). *Scrutinizing systemic risks in personalized recommender systems through sock-puppet auditing of VLOPs*. ACM Transactions on Recommender Systems. <https://doi.org/10.1145/3795516> (auditing); Kollnig, K., Datta, S., & Van Kleek, M. (2021). *I want my app that way: Reclaiming sovereignty over personal devices*. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems (CHI EA '21)*. ACM. <https://doi.org/10.1145/3411763.3451632> (user autonomy); Palmieri, A., Kollnig, K., Tamò-Larrioux, A. (2026). *Systemic risks of dominant online platforms: A scoping review*. Computer Law & Security Review. <https://doi.org/10.1016/j.clsr.2026.106262> (systemic risks).

<sup>3</sup>

<https://www.reuters.com/business/google-apple-enter-into-multi-year-ai-deal-gemini-models-2026-01-12/>

<sup>4</sup> [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4780718](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4780718)

<sup>5</sup> <https://searchengineland.com/openai-chatgpt-serpapi-google-search-results-461226>

To some extent, this is not wholly surprising, given that it is extremely difficult and expensive to build up a competitor in information discovery and access.

Further, regarding attribution, we would emphasise that there is a need for Publishers and the CMA to understand how decisions around attribution are made, not just how content is attributed. On this basis, Google's published explanation should include elements such as a description of how sources are selected and ranked for inclusion in a generated response, an explanation of how attribution is displayed to users and a description of how Google monitors attribution accuracy.

On the second element, it would also be helpful for further information to be provided on if and how the display format of the explanation provided to users on different device types, query categories, or user segments differ in ways that could also contribute to affecting publisher outcomes.

We also strongly support making this information publicly available, rather than providing it only to registered publishers via the Search Console. The question of how AI systems use and represent web content is a matter of public interest, not only a matter of commercial concern to publishers. In line with our existing work on [researcher access to platform data](#), we would emphasise the need for this data to be made publicly available, such that independent researchers and civil society organisations can be empowered to assess Google's claimed practices against observable outputs.

**6.4(c) We welcome further views on the extent to which our proposed Publisher CR can be expected to result in the identified consumer benefits, including ensuring that users are able to assess and trust the content they read on the web**

The proposed Publisher CR is a step towards the identified consumer benefits, however the extent to which it will achieve its intended aim will depend heavily on how it is implemented.

The primary consumer benefit identified concentrates on enabling users to assess and trust AI-generated content. This is a question of information quality and provenance, which has also received attention in recent [research from Ofcom on answer engines](#), in which they raise concerns around the transparency and attribution of sources. Ultimately, users need to be able to identify the source of information to be able to sufficiently assess the credibility of the provided information. They also need to be able to access underlying content if they look to verify any claims included within the answer they have been provided. Trust can only be developed if this is possible, which therefore poses risks to the realisation of the intended consumer benefits proposed by the CMA.

Even with improved attribution, users can only verify AI-generated content if the attributed source links lead them directly to content that substantiates the claim made in the

summary. At present, where Google's generative AI features paraphrase publisher content, accurate attribution to the source page does not currently guarantee that the user will find the verification they are looking for. This is a structural limitation of the current approach that the CMA should acknowledge.

Another risk to the realisation of the anticipated benefit from the Publisher CR is that these shall likely be unevenly distributed. Users with higher data and digital literacy will be better placed to make use of source links and to assess the credibility and veracity of cited sources, whereas others may struggle. The CMA should encourage Google to consider how attribution is designed for users with such varying levels of data literacy, such as whether source identifiers are presented in ways that are meaningful to users unfamiliar with particular publishers.

## Data Portability

Implemented correctly, data portability promotes a resilient, competitive, and most importantly open data ecosystem. We champion not only portable data, but strong investment in the infrastructure necessary to ensure that data sources are secure, available, and interoperable for data subjects' personal access, research, and to support innovative markets.<sup>6</sup>

### ***(b) Do you consider the proposed Data Portability CR would result in the potential benefits we have identified (for example, value and innovation)?***

Reliable and predictable data portability is critical for fostering competition, but should also ultimately benefit data subjects. We foresee the Data Portability CR as an important mechanism for individuals to have greater personal agency over their data held by Google. Storage of data in personal data stores like [Solid pods](#), for instance, allows data subjects low-level control over their data in a secure environment.

Enabling users to port their data securely and reliably from Google into their Solid pod would open opportunities for data subjects to share their data forward on a one-off basis without losing control of it. To this end, we encourage that such personal data storage services be considered when establishing requirements for security vetting. Some Solid server providers such as individuals, universities or civil society organisations may be able to gain CASA certification, but others may not be able to afford its cost.

Furthermore, [data donations for the purpose of research](#) may also be a legitimate purpose that users request data portability. For these cases, alternative security requirements like those required of controllers under GDPR should be considered. [UK GDPR Article 20](#) also

---

<sup>6</sup> See our responses on [Open Banking](#) and [Open Communications](#)

grants individuals data portability rights, which ideally could be supported by a similar (or the very same) infrastructure and should follow an interoperable data standard.

An important shortcoming of the CR's data portability requirement is the lack of real-time transfer of data, as previously in an analysis of the shortcomings of data portability by CoCoDa researchers.<sup>7</sup> This makes it difficult to build innovative new businesses that wish to compete with and innovate beyond the status quo of Google as a dominant market player.

***(d) Do you agree with the content of the Interpretative Notes? Are they sufficiently clear and comprehensive? Do they cover the right issues? Are there any gaps?***

We welcome the interpretive notes and their acknowledgement of the need for greater technical specification beyond only legal mandate. Moving forward, we see them as an important tool to continue supporting interoperability with EU DMA requirements and other international access provisions as they evolve.

We encourage that future notes be used when further data standards are published both under the DMA and the [UK Smart Data Initiative](#). Interpretation and restructuring of data can be burdensome and costly for API users.<sup>8</sup> When fields, response codes, or other elements of data structure change without prior notification, downstream systems can be adversely affected.<sup>9</sup> Keeping up with structures will be key in delivering the benefits intended in the CR.

One area which is not covered is real-time time access to data, as is guaranteed in the DMA. For a business to be built around access to data via the portability API, data will often need to move at near or [“functional real time”](#).

In the future, the CMA could use Interpretative Notes for:

1. Aligning to DMA or UK Smart Data Standards.
2. Requiring detailed data catalogues and specifications that include both plain language descriptions of data contents, how data is produced, and how often it is refreshed, as well as raw API query parameters and query result formats.

---

<sup>7</sup> <https://www.sciencedirect.com/science/article/abs/pii/S2212473X25000082?via%3Dihub>

<sup>8</sup> Bohr, J. Coordination transparency: governing distributed agency in AI systems. *AI & Soc* (2026). <https://doi.org/10.1007/s00146-026-02853-w>

<sup>9</sup> Binns, R., Stein, J., Datta, S., Van Kleek, M. and Shadbolt, N., 2025, June. Not Even Nice Work If You Can Get It; A Longitudinal Study of Uber's Algorithmic Pay and Pricing. In *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency* (pp. 1484-1497).

3. Clarifying advance notification requirements and procedures for changes to API parameters and result formats.
4. Establishing standards and assurances about latency of [real time data](#).

These uses would ensure forward-compatibility, while considering that legal requirements alone are insufficient without stable, transparent and interoperable technical infrastructures that lower barriers to entry.

**(e) Do you agree with our proposals for compliance reporting and for monitoring the effectiveness of the proposed intervention? Have we identified the right metrics?**

Without an SLA, reporting and monitoring will be an essential tool if the data portability imagined in the CR is to be successful. Annual reporting will be insufficient in assuring the quality of data used on this timescale. With the DTI, ODI researcher Thomas Carey-Wilson recently produced a regulator's "[toolkit for assessing data portability under the DMA and digital competition laws](#)".

The EU commission has produced [a set of dashboards](#) that refresh on a 24-hour cycle as part of its Digital Services Act (DSA). Critical services of Google's API's scale require this level of reporting to sufficiently inform services that may be reliant on data portability infrastructure. Such reporting tools should be considered as API uptake grows.