Reclaim the Web!

Policy Paper

Boosting digital sovereignty and competitiveness through a European Web Data Infrastructure

Edited and published by the European Open Web Search Initiative

July 2026

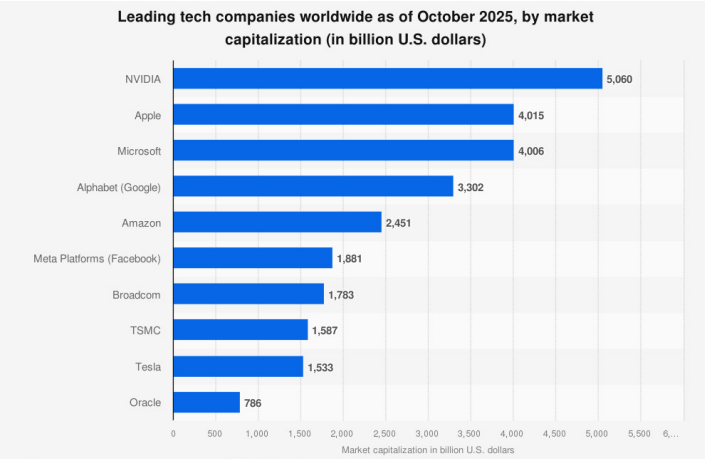
Open WebSearch .eu

Boosting digital sovereignty and competitiveness through a European Web Data Infrastructure

Summary

Europe must build a Web Data Infrastructure that provides European start-ups, SMEs and industry with large-scale access to high-quality Web data.
Digital services and business models that are heavily used by European consumers, businesses and authorities need large amounts of Web-based text, images, and videos. These include search, AI, business intelligence, shopping, streaming, social media, and many more.

Due to the complexity and the costs involved, large-scale and systematic access to Web data is today mainly controlled by US-based Big Tech companies. Most European companies lack the resources to collect the large volumes of Web data they would need to build competitive digital services by themselves.



source: https://www.statista.com/statistics/1350976/leading-tech-companies-worldwide-by-market-cap/

Eight American and one Taiwanese company own the world's main information infrastructure.

What does this mean for Europe?

The complexity of European regulations, particularly with respect to copyright, data protection and liability, poses another barrier for market entry. This leaves Europe dependent on digital services from overseas that in many cases could not even be built in Europe in accordance with current European legislation.
This critical dependency on overseas tech companies is disastrous for Europe's economy, democracy and rule of law. It results in a dramatic loss in domestic European value creation and facilitates manipulation through disinformation as well as black-box algorithms. It also drastically hinders Europe from enforcing its digital regulation.

In order to break up this vicious circle and to boost digital sovereignty and competitiveness, Europe must implement the following **core demands:**

- **build and operate a pan-European Web Data Infrastructure**
- **partially adapt the regulatory framework**
- **allocate respective base funding in the upcoming Multiannual Financial Framework / European Competitiveness Fund as well as in Horizon Europe and FP10**

➡ In case storage and compute would be provided via existing public data centres, **required funds** to operate a fully established European Web Data Infrastructure are estimated to be around: **~50M€ p.a.**

Don't let Europe get unplugged!

Europe must act fast and boldly and put in place the organisational, financial and legal conditions for the development of new, diverse European Web services and thus the establishment of digital sovereignty and competitiveness.

The foundation has already been laid – we must now build on it.

Europe: The digitally dependent continent

Europe lacks digital sovereignty. It is largely dependent on Web services from overseas and, under current economic and legal circumstances, unable to break free. Today mostly US based Big Tech companies have established very successful, partially monopolistic, services and business models in the Web that are indispensable for consumers, businesses and the authorities today.
These services build on the large-scale commercial use of web-based text, images, and video, here referred to as 'Web data'. They include search, AI, business intelligence, shopping, streaming, social media, and many more.

Web data is one of the essential resources for building digital services and businesses.

In contrast to the overseas hyper-scalers, European digital industry, start-ups, and innovators face severe market entry barriers that especially SMEs cannot clear.
On the economic side, most European companies are not in the position to collect the data they need to build innovative Web services by themselves. Doing so would require constant crawling of the Web, preprocessing and indexing of the data collected - which, due to the sheer scale of information on the Web, is very compute-intensive and costly.
On the legal side, the complexity of European regulation, particularly with respect to copyright, data protection and liability, poses an additional barrier for market entry.

The disastrous consequences of digital dependency

The lack of digital sovereignty is disastrous for Europe's economy, democracy and rule of law.
From an economic perspective, the dependence on overseas Big Tech companies undermines the competitiveness of European businesses and industries dramatically. Europe thus loses out significantly on domestic value creation in the digital domain and keeps falling behind economically.
At the same time, the imbalance of power in the digital space jeopardizes Europe's democracy. Public opinion and the political process are open to manipulation through disinformation and black-box algorithms.
On the legal side, overseas Big Tech companies routinely violate European laws and regulations regarding, for example, privacy, competition law and copyright. Europe's dependence on these companies prevents it from enacting and effectively enforcing applicable rules and regulations.

Why the digital economy heavily relies on Web data

The digital economy runs on large-scale commercial use of Web data (text, images, video).

High-quality Web data is essential for most digital services and innovations:

- Large-language models like, for example, ChatGPT (OpenAI) or Gemini (Google) are trained on massive amounts of Web data
- Any type of Web search and information access, whether through classical search engines or AI chatbots, requires a structured and constantly updated index of information and content available on the Web
- Online maps need to be enriched with up-to-date information like opening hours, contact details, and FAQs
- Business analytics, market statistics as well as strategic information and many more

➡ The raw data for all of these needs to be retrieved from the Internet. The lack of large-scale access to Web data and legal uncertainty around its use cripples Europe's digital economy and innovation.

➡ **Europe must act now and build its own Web Data Infrastructure!**

The solution: A European Web Data Infrastructure

To reduce its dependency from overseas Big Tech companies and to boost digital sovereignty and competitiveness, Europe needs its own Web Data Infrastructure (including an Open Web Index) and the regulatory framework for its creation and use. This will significantly lower the market entry barrier and will enable the building and operation of competitive and compliant domestic European Web services.[1]

A European Web Data Infrastructure will offer large-scale access to Web data for all European companies, SMEs and innovators.
The proof of concept is already available.

Europe has started to build such a publicly accessible Web Data Infrastructure in the Horizon Europe-funded OpenWebSearch.EU[2] project. The pilot is available via openwebindex.eu[3] and offers structured and filterable Web data, extended daily. It is built on principles of trust, transparency and openness and is available for testing under research conditions. Europe must build on this proof-of-concept and implement, scale, and operationalise this infrastructure. With a European Web Data Infrastructure, Europe can build its own diverse and competitive AI, search and Web analytical services and boost its digital sovereignty and competitiveness, similar to how it does with its Copernicus[4] and Galileo[5] programmes for monitoring and navigation of the physical Earth.

Organisational aspects

Europe must now establish the necessary financial and organisational structures to build and operate a fully operational European Web Data Infrastructure.
The best place to host and operate this would be the existing ecosystem of High Performance Computing Centres (EuroHPC)[6], AI Factories and Antennas. This way, the European Web Data Infrastructure could make resource- and cost-efficient use of existing supercomputing infrastructure (storage and compute) and support the mission of the AI Factories with high-quality Web data.
Europe must establish dedicated base funding in the upcoming Multiannual Financial Framework (MFF)/European Competitiveness Fund ensuring sufficient and long-term financial resources for building and operating the European Web Data Infrastructure. This has to be accompanied by continuous related research and development work on big data, Web and retrieval technologies through Horizon Europe and FP10.

[1] https://openwebsearch.eu/market-potential-study
[2] https://openwebsearch.eu/
[3] https://openwebindex.eu/
[4] https://www.copernicus.eu/en
[5] https://www.euspa.europa.eu/eu-space-programme/galileo
[6] https://www.eurohpc-ju.europa.eu

Costs for operating a fully established European Web Data Infrastructure with hubs in all EU member states are estimated to be on the order of 200M€ p.a.[7]

In case storage and compute for the European Web Data Infrastructure would be provided via existing public data centres, funds required to operate the Infrastructure are estimated to be around ~50M€ p.a. Not all funds necessarily have to come from the public domain, since parts of this infrastructure could also be established and operated in a public-private-partnership.

Regulatory aspects

In addition to financing and setting up such an infrastructure technically, the legal basis for creating and operating the European Web Data Infrastructure and its use by start-ups, digital innovators, industry, researchers, and SMEs in Europe also must be established. This involves all elements of data generation and handling, in particular the crawling, preprocessing, indexing and sharing of Web data via a public infrastructure and also for commercial use. Current European regulatory frameworks pose risks to both the operator of a European Web Data Infrastructure as well as to its commercial users. Key legal aspects that need clarification are related to publicly available Web content being systematically downloaded, processed and shared through the European Web Data Infrastructure (in a bulk manner) and later on being used for commercial purposes. The Web content being collected and shared includes personal information, copyrighted content, and possibly illegal material.

When collecting Web data and incorporating it in their services, large overseas technology companies regularly violate applicable European law and factor in the legal costs and fines associated with possible violations. At the same time, the lack of sovereignty makes it difficult for Europe to effectively enforce its digital and privacy regulation. To level the playing field, Europe needs to create the legal framework for building and operating a public Web Data Infrastructure as well as its commercial use.

An important prerequisite for the applicability of the following regulatory adaptations will be the adoption of standards and auditing procedures that the European Web Data Infrastructure has to adhere to. This will include transparency and reporting over the included/stored websites, clearly defined procedures for removal of content on request, as well as systematic external and public auditing to ensure neutrality and to prevent misuse. Fairness and accessibility must be achieved through openness. The design of the European Web Data Infrastructure must ensure that fundamental rights are respected and privacy is protected.

[7] https://www.openwebsearch.eu/market-potential-study

Regulatory clarifications are required mainly in three domains:

1) Explicitly allowing processing and redistributing of publicly accessible Web content for the purpose of creating and maintaining a European Web Data Infrastructure in the public interest by means of a statutory exception or extension of the Digital Single Market Directive (2019/790).

2) Clarifying that the European Web Data Infrastructure falls under the liability exemptions for intermediary services of the DSA (2022/2065). At the same time, making the crawled data available to third parties must legally be treated as a neutral data intermediary activity, excluding liability for content, provided the intermediary service acts with proportional due diligence and does not act unlawfully by intent.

3) Including a specific legal basis under Art. 6(1)(e) GDPR (2016/679) ("task carried out in the public interest") and Art. 9(2)(g) GDPR for the collection and redistribution of publicly available Web data by the European Web Data Infrastructure. This of course requires mechanisms for safeguards and technical data minimisation.

In addition, the users of the infrastructure need legal certainty for how to build innovative digital services and European alternatives to the dominant overseas players in a legally compliant way.
This requires sound balancing of interests of application developers and Web service providers on the one hand and the content creators and data subjects on the other. One factor in balancing these interests could be the implementation of a micro-payment scheme for reimbursement between the providers of Web content (journalism, news, arts, creation, etc.) and commercial users thereof (AI systems, online market places, commercial search engines etc.).

Core regulatory aspects

- **DSM Copyright Directive:** Allow processing and redistributing of publicly accessible Web content for creating and maintaining the European Web Data Infrastructure
- **DSA:** Include the European Web Data Infrastructure into the liability exemptions for intermediary services
- **GDPR:** Clarify the legal basis for the collection and redistribution of publicly available Web data by the European Web Data Infrastructure

As prerequisite, a European Web Data Infrastructure must adhere to:

- standards and auditing procedures
- transparency, neutrality, fairness, openness, accessibility
- safeguarding fundamental rights and privacy

Outlook and way ahead

Providing Web data through a European Web Data Infrastructure, with secured long-term base funding as well as legal certainty in the usage of Web data, will significantly boost European innovation competitiveness and value creation in the digital sphere and beyond.

A European Web Data Infrastructure, accessible for all European actors, enables the development and operation of many domestic European digital services that can compete with overseas Big Tech offerings.

Growing this ecosystem of legally compliant European digital services will empower the European Commission and Member States to further enforce its digital regulations against the tech giants.

Fostering European alternative digital services by providing a European Web Data Infrastructure helps also to overcome the paradox situation that Europe relies on overseas digital services that could in many cases not even be built domestically and in accordance with current European laws, resulting in a significant loss in European value creation and missing tech sovereignty.

Act now What should Europe do next?

1. Set up the Web Data Infrastructure on a solid operational basis

Why? A sustainable infrastructure that represents European diversity and serves the economy and citizens is indispensable to strengthen democracy and economic growth.

2. Ensure legal clarity regarding a European Web Data Infrastructure

Why? The current ground-base has been built and needs to be maintained and further developed to be able to compete with overseas

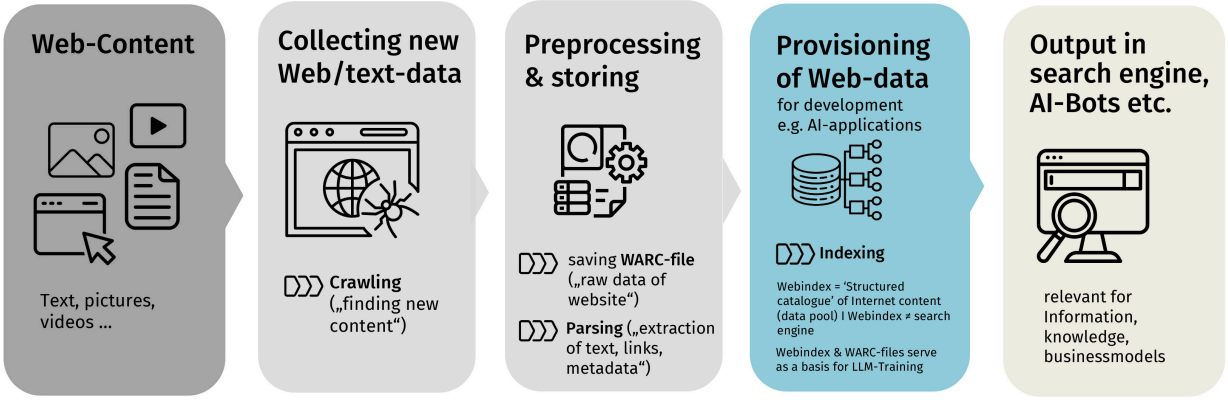
3. Ensure long-term funding in the upcoming MFF/ECF and FP10

Why? The foundations and a prototype have been built in OpenWebSearch.EU. Now this needs to be further developed, operationalized and scaled in order to support the building of domestic and competitive European Web services.

Interested in how it works technically?

Back-Up Information on a Web Data Infrastructure

Web Data Infrastructure - the process behind (simplified representation)



Contact:

Open Search Foundation e.V.
The Open Search Foundation e.V. - www.opensearchfoundation.org - is a non-profit organisation registered in Germany and serves as network coordinator and contact point for the European Open Web Search Initiative.

ows@openwebsearch.eu

www.openwebsearch.eu