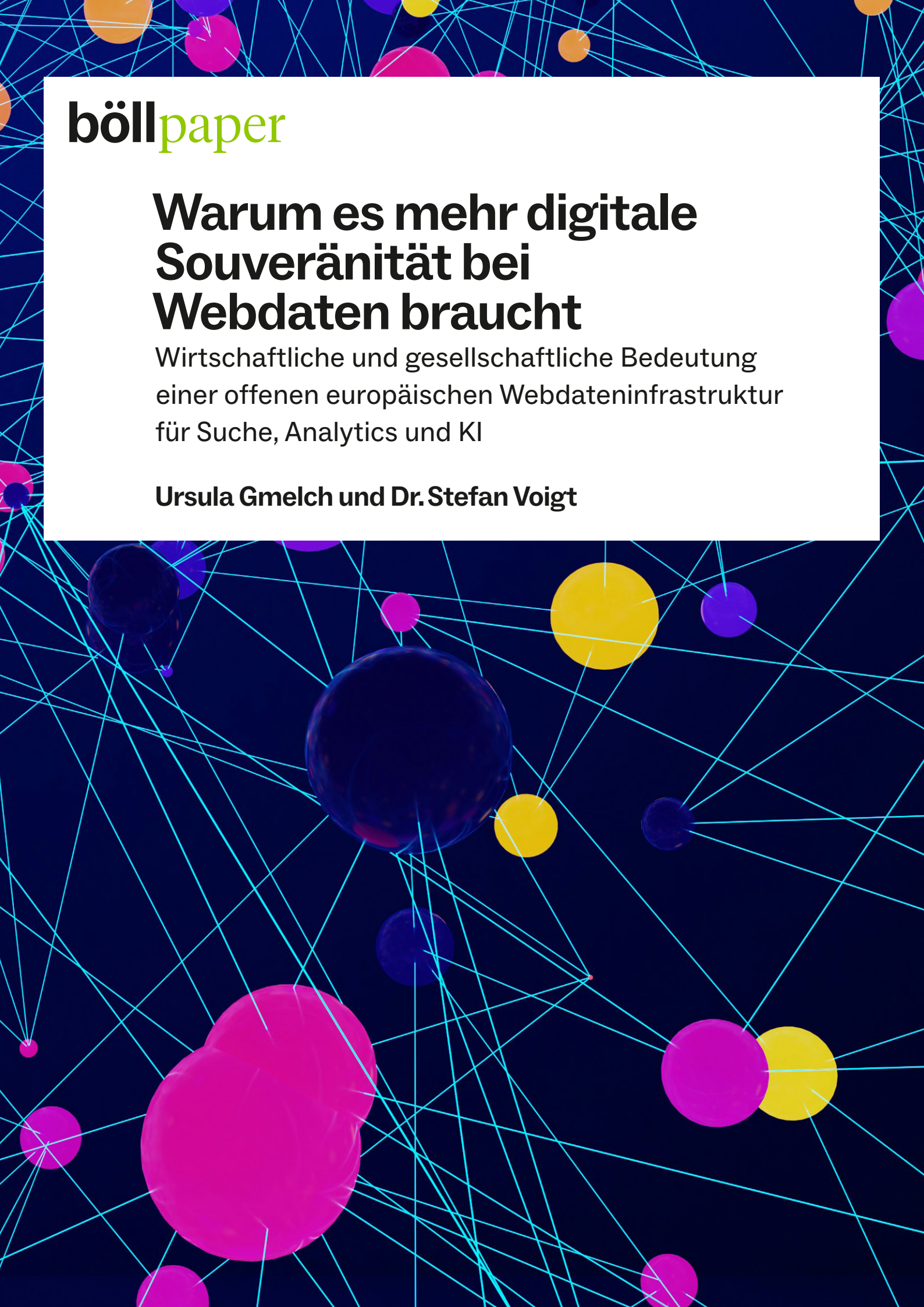
böllpaper

Warum es mehr digitale Souveränität bei Webdaten braucht

Wirtschaftliche und gesellschaftliche Bedeutung einer offenen europäischen Webdateninfrastruktur für Suche, Analytics und KI

Ursula Gmelch und Dr. Stefan Voigt



Warum es mehr digitale Souveränität bei Webdaten braucht

Wirtschaftliche und gesellschaftliche Bedeutung einer offenen europäischen Webdateninfrastruktur für Suche, Analytics und KI

Ein Policy Paper von Ursula Gmelch und Dr. Stefan Voigt

HEINRICH BÖLL STIFTUNG

Inhalt

Zusammenfassung 4

1 Deutschland ist bei der Websuche und bei Webdaten für KI-Systeme digital abhängig 5

2 Warum ist Deutschlands digitale Abhängigkeit ein Problem? 7

3 Essenzieller Baustein digitaler Souveränität: eine europäische Webdateninfrastruktur für eine Vielzahl von Diensten und Anwendungen 8

4 Gesellschaftliches und wirtschaftliches Potenzial einer europäischen Webdateninfrastruktur 10

Digitale Souveränität 10
Wirtschaftliche Vorteile einer europäischen Webdateninfrastruktur 10
Soziale und kulturelle Auswirkungen 11
Umwelt- und Sicherheitsaspekte 11

5 Aufbau und Betrieb einer europäischen Webdateninfrastruktur: organisatorische, finanzielle und regulatorische Aspekte 13

Organisatorische Rahmenbedingungen 13
Finanzierung – Bedarf und Potenziale 14
Regulatorische Anpassungen 15

6 Fazit und Handlungsempfehlungen 17

Die Autor*innen 18
Impressum 19

Zusammenfassung

Deutschland und Europa sind digital abhängig, auch bei der Websuche sowie hinsichtlich Webdaten für KI-Systeme. Diese Abhängigkeit stellt sowohl ein wirtschaftliches Problem als auch eine Bedrohung für europäische Werte und die demokratische Ordnung dar.

Eine europäische Webdateninfrastruktur ermöglicht die Entwicklung und den Betrieb zahlreicher deutscher und europäischer digitaler Dienste, die mit den Angeboten der großen US-amerikanischen Technologiekonzerne konkurrieren können. Deutschland muss sich dringend und verstärkt in Europa für eine solche öffentliche europäische Webdateninfrastruktur einsetzen.

1 Deutschland ist bei der Websuche und bei Webdaten für KI-Systeme digital abhängig

Deutschland und Europa sind digital abhängig von ausländischen Technologieunternehmen, vor allem von US-amerikanischen Big-Tech-Konzernen. Das Problembewusstsein hierfür wächst in politischen Kreisen und zunehmend auch bei den Bürger*innen (siehe etwa die Digital-Independence-Day-Kampagne, bei der viele Organisationen gemeinsam für den Wechsel zu demokratiefreundlichen digitalen Alternativen werben). Es wird zunehmend erkannt und diskutiert, dass die digitalen Abhängigkeiten für Wirtschaft und Gesellschaft kritische Konsequenzen haben: einen dramatischen Verlust an Wertschöpfung in Deutschland und Europa, Manipulation durch Desinformation und Black-Box-Algorithmen, eine stark erschwerte Durchsetzung europäischer Digitalregulierung und viele mehr.

Eine Lösung für diese Abhängigkeiten sind europäische digitale Infrastrukturen. Die EuroStack-Initiative hat hierfür eine Vision in den politischen Mainstream gebracht. Sie setzt sich dafür ein, auf allen Ebenen des Stacks (Gesamtheit der Technologien, die bei der Erstellung einer Softwareanwendung zusammenarbeiten) europäische Alternativen zu fördern.

Ein Aspekt, der hinsichtlich digitaler Abhängigkeit jedoch noch viel zu wenig Aufmerksamkeit erfährt, ist der souveräne Zugang zu Informationen bzw. zu Daten im Web. Für unzählige digitale Produkte und Dienstleistungen – ob Suchmaschinen, KI-Anwendungen, Business Analytics, E-Commerce, Streaming oder soziale Medien – sind Daten und insbesondere Webdaten ein unentbehrlicher Rohstoff. Dies betrifft Information in Textform, aber auch Video- und Bildformate. Auch hier sind Deutschland und Europa weitgehend abhängig von Big Tech aus Übersee.

Ein Blick in den Suchmarkt – der am stärksten vom Zugang zu Webdaten im großen Stil abhängt – zeigt ein verheerendes Bild: Bei den klassischen Suchmaschinen hatte Google im April 2026 in Deutschland einen Marktanteil von über 80 Prozent, weit vor Microsoft Bing mit nicht einmal 10 Prozent. Bereits 2023 hat die Europäische Kommission Googles Mutterfirma Alphabet im Rahmen des Digital Markets Act als Torwächter (Gatekeeper) eingestuft. Als Gatekeeper gelten Unternehmen, die über zentrale Plattformdienste ein wichtiges Zugangstor zu Verbraucher*innen kontrollieren und damit große Marktmacht haben. Auch alternative Suchmaschinen wie Ecosia, Qwant oder DuckDuckGo greifen auf die Ergebnisse und Rankings von Google oder Microsoft Bing zurück. Ecosia hat sich inzwischen mit Qwant in einem Joint Venture zusammengeschlossen, um zumindest einen Teil der Suchergebnisse mittelfristig aus einer eigenen Infrastruktur und einem eigenen Webindex liefern zu können.

In den letzten Jahren ist zu der klassischen Suchmaschine die KI-gestützte Suche hinzugekommen, etwa wenn ChatGPT oder Claude zur Beantwortung einer Anfrage auf aktuelle Informationen aus dem Internet zugreifen. Auch bei vielen klassischen Suchmaschinen wird inzwischen zuerst eine KI-generierte Zusammenfassung der Suchergebnisse angezeigt (z. B. Google AI Overviews). Dadurch verschwimmen die Grenzen zwischen Suche und KI-Systemen zunehmend.

Die Suche im Internet verändert sich also, sie wird deshalb jedoch keineswegs obsolet – im Gegenteil: Auch KI-Systeme brauchen für die Generierung von Antworten verlässliche und aktuelle Informationen aus dem Web. Und sie bleibt von US-amerikanischen Unternehmen dominiert, die damit kontrollieren, welche Informationen Europäer*innen sehen.

Deutschland und Europa brauchen daher dringend einen unabhängigen, souveränen und umfassenden Zugang zu Informationen und Datenbeständen im Web. Hierzu bedarf es einer souveränen europäischen Webdateninfrastruktur.

2 Warum ist Deutschlands digitale Abhängigkeit ein Problem?

Deutschlands und Europas digitale und informationelle Abhängigkeit stellt sowohl ein wirtschaftliches Problem als auch eine Bedrohung für europäische Werte und die demokratische Ordnung dar.

Wirtschaftlich untergräbt die Abhängigkeit von außereuropäischen Big-Tech-Firmen die Wettbewerbsfähigkeit deutscher Unternehmen. Deutschland verliert dadurch im digitalen Bereich erheblich an inländischer Wertschöpfung und fällt wirtschaftlich immer weiter zurück. So lag das Leistungsdefizit der EU zu den USA bei digital erbringbaren Dienstleistungen 2023 bei 145 Milliarden Euro, 2024 sogar bei 173 Milliarden Euro. Dieser Umsatz entgeht der europäischen Wirtschaft.

Gleichzeitig gefährdet das Machtungleichgewicht im digitalen Raum die Demokratie in Deutschland und Europa. Öffentliche Meinung, Wahlen und politische Prozesse sind durch Desinformation und undurchsichtige Algorithmen anfällig für Manipulationen.

In rechtlicher Hinsicht verstoßen ausländische Big-Tech-Unternehmen regelmäßig gegen deutsche und europäische Gesetze und Vorschriften, beispielsweise in den Bereichen Datenschutz, Wettbewerbsrecht und Urheberrecht. Die Abhängigkeit Deutschlands und Europas von diesen Unternehmen hindert uns jedoch in vielen Fällen daran, nötige Regeln und Vorschriften zum Schutz unserer Demokratie und Wirtschaft zu erlassen und geltende Gesetze wirksam durchzusetzen.

3 Essenzieller Baustein digitaler Souveränität: eine europäische Webdateninfrastruktur für eine Vielzahl von Diensten und Anwendungen

Bei der Entwicklung von datenintensiven Webdiensten sind deutsche und europäische Digitalunternehmen, Start-ups und Innovator*innen immensen Markteintrittshürden ausgesetzt, die insbesondere für kleine und mittlere Unternehmen (KMU) und den digitalen Mittelstand schwer zu überwinden sind. Eine dieser Markteintrittshürden besteht darin, die umfassenden Webdaten, die für vielfältige innovative Anwendungen und Dienstleistungen benötigt werden, zu beschaffen und rechtskonform zu verarbeiten. Hierzu ist es nötig, umfassend und dauerhaft das Web zu crawlen, die gesammelten Webdaten zu verarbeiten, zu filtern, zu indexieren und für Analytics-Dienste, Suchmaschinen oder agentenbasierte KI-Systeme verwendbar zu machen. Da das Web riesig ist, ist der Aufwand enorm – sowohl finanziell als auch hinsichtlich der benötigten Rechen- und Speicherkapazitäten. Dies führt aktuell dazu, dass sich deutsche und europäische Digitalunternehmen, Start-ups und Innovator*innen von den Datenschnittstellen außereuropäischer Big-Tech-Konzerne (vor allem Google und Microsoft) abhängig machen müssen.

Bei den webdatenbasierten Anwendungen geht es ausdrücklich nicht nur um Suchmaschinen – ob klassisch oder als KI-Chatbots –, sondern um eine Vielzahl an digitalen Dienstleistungen. So brauchen etwa digitale Kartendienste umfangreiche, aktuelle und verlässliche Point-of-Interest-Informationen wie Öffnungszeiten, Kontaktdaten, Ortsangaben und FAQs. Statistische Ämter auf Landes-, Bundes- und europäischer Ebene oder Business-Analytics-Dienstleister können vom Zugang zu Webinformationen im großen Maßstab profitieren. Auch in der Forschung gibt es unzählige Anwendungsbereiche, etwa die Ergänzung der physikalischen Erdbeobachtung durch webbasierte Rauminformation. Ein weiteres Beispiel sind die Erstellung und der Betrieb von souveränen und sicherheitsrelevanten agentenbasierten Fach-KI-Systemen in Firmen oder Behörden.

All diese digitalen Dienste und Anwendungen können in Deutschland und Europa nur dann souverän entwickelt und betrieben werden, wenn eine eigene öffentliche Webdateninfrastruktur zur Verfügung steht. Dazu gehört auch ein offener Webindex: ein strukturiertes Abbild des Webs und der Kern jeder Suchmaschine. Für eine solche Webdateninfrastruktur wurde bereits ein Konzept entwickelt: Zuerst wird das öffentlich zugängliche Internet tagesaktuell gecrawlt. Die gesammelten Daten und Informationen werden anschließend verarbeitet, gefiltert, indexiert und europäischen Forschenden, Innovator*innen und Digitalunternehmen für den Aufbau von

Mehrwertdiensten im Bereich Websuche, Analytics und KI-Dienstleistungen etc. zur Verfügung gestellt. Die Webdateninfrastruktur kann zudem als Mittlerin zwischen Urheber*innen und Nutzenden digitaler Inhalte dienen und so Mikrozahlungen an Kunstschaffende, Journalist*innen, Verlage usw. ermöglichen.

Aufbau und Betrieb der Webdateninfrastruktur basieren auf drei Grundprinzipien:

- **Kooperatives, verteiltes Rechnen:** Die Infrastruktur wird ressourcenschonend und unter Nutzung bestehender öffentlicher und privater Rechenkapazitäten in ganz Europa aufgebaut.
- **Open-Source-Algorithmen:** Der Einsatz von Open-Source-Technologien gewährleistet die Transparenz und Integrität des Gesamtsystems als öffentliche Infrastruktur.
- **Öffentliche und transparente Moderation und Kuratierung:** Struktur und Inhalte der Webdateninfrastruktur werden kontinuierlich in Kooperation von Wissenschaft, Bürgervertreter*innen und Gremien in den jeweiligen EU-Mitgliedsstaaten auditiert. So kann diese wichtige Webdateninfrastruktur offen, transparent und vertrauenswürdig betrieben und genutzt werden.

Der Prototyp einer solchen europäischen Webdateninfrastruktur, einschließlich eines Open Web Index, wurde im OpenWebSearch.EU-Projekt (gefördert von 2022 bis 2026 im Rahmen von Horizon Europe) erstellt und erprobt. Jetzt hat Europa die Chance, auf diesem Proof-of-Concept aufzubauen, es zu skalieren und in den EU-Mitgliedsstaaten gemeinschaftlich zu implementieren. Hierfür ist die aktive Unterstützung von staatlicher Seite (sowohl in Deutschland als auch in den anderen EU-Mitgliedsstaaten und auf europäischer Ebene) unabdinglich.

Auch die European Search Perspective – das Joint Venture von Qwant und Ecosia – argumentiert, dass die Suchinfrastruktur als strategische öffentliche Infrastruktur behandelt werden sollte. Der Zugang zu Informationen sowie zu digitalen Märkten ist ein Element der Daseinsvorsorge und darf nicht einigen wenigen kommerziell orientierten Gatekeepern überlassen werden. Neben einer aktiven staatlichen Unterstützung sind auch Public-Private-Partnership-Modelle für den Betrieb einer Webdateninfrastruktur denkbar.

4 Gesellschaftliches und wirtschaftliches Potenzial einer europäischen Webdateninfrastruktur

Digitale Souveränität

Die Einrichtung einer europäischen Webdateninfrastruktur bedeutet einen signifikanten Gewinn an digitaler Souveränität. Sie ist eine autonome digitale Infrastruktur für Europa, die unabhängig von nichteuropäischen Anbieter*innen betrieben werden kann. Deutschland und Europa profitieren davon durch neue eigene Digitaldienste, reduzierte Vendor-Lock-in-Effekte (leichterer Wechsel des Anbieters) und mehr Wahlmöglichkeiten für Unternehmen und Verbraucher*innen.

Wie bereits in Kapitel 3 erläutert, senkt eine europäische Webdateninfrastruktur die Markteintrittshürden insbesondere für den digitalen Mittelstand, indem sie deutschen und europäischen Digitalunternehmen und Start-ups für die Entwicklung von Digitaldiensten Webdaten in großem Umfang und gemäß europäischem Recht bereitstellt – ohne dass sie hierfür jeweils selbst eine eigene Infrastruktur aufwendig aufbauen müssen. Damit können in Deutschland und Europa vielfältige neue digitale Dienste entwickelt werden. Mehr Vielfalt in den Dienstleistungen bedeutet mehr Ausfallsicherheit und weniger Abhängigkeit von einigen wenigen Hyperscalern.

Das Mehr an Wahlmöglichkeiten für Nutzer*innen und Unternehmen erlaubt es ihnen, ihre jeweiligen Präferenzen besser durchzusetzen – zum Beispiel die Nutzung digitaler Dienste und Plattformen, die den europäischen Datenschutz- und Urheberrechtsgesetzen entsprechen oder anderweitig europäische Werte besser umsetzen als die bestehenden Plattformen der Big-Tech-Anbieter.

Wirtschaftliche Vorteile einer europäischen Webdateninfrastruktur

Mehr digitale Souveränität sorgt dafür, dass nichteuropäische Gatekeeper nicht mehr so viel digitale Wertschöpfung von deutschen und europäischen Internetnutzenden abziehen. Weniger Gebühren für die Nutzung nichteuropäischer digitaler Dienste fließen ins Ausland ab, vielmehr bleibt ein Teil der Wertschöpfung in Deutschland und Europa, wodurch der Wettbewerb im digitalen Raum ausgewogener wird bzw. erst wieder entstehen kann.

Eine europäische Webdateninfrastruktur trägt dazu bei, indem sie (Daten-)Markteintrittsbarrieren reduziert und dadurch Innovation, europäische Start-ups und Open-Source-Entwicklung fördert. Zudem stärkt sie Deutschlands und Europas Position im globalen Technologie-, IT- und Datenmarkt und ermöglicht eine verbesserte Datenverfügbarkeit und -verwertung in Schlüsselindustrien.

Über einen Zeitraum von zehn Jahren werden die positiven Auswirkungen einer europäischen Webdateninfrastruktur auf mehr als 5,6 Milliarden Euro geschätzt, mit einem Nettonutzen (abzüglich der Aufbau- und Betriebskosten) von über 4 Milliarden Euro.

Soziale und kulturelle Auswirkungen

Neben dem Souveränitätsgewinn und den wirtschaftlichen Chancen kann die europäische Webdateninfrastruktur auch eine positive soziale, kulturelle und demokratisierende Wirkung entfalten. So wird der Zugang zu digitalem Wissen und zu Online-Informationen demokratisiert, ohne dass sich Deutschland und Europa von außereuropäischen und kommerziell getriebenen Anbieter*innen abhängig machen müssen. Open Science (Offene Wissenschaft) und Forschung werden gestärkt, da mit dem Web eine riesige Informationsquelle systematisch und im großen Maßstab nutzbar gemacht und so der Zugang zu webbasierter Forschungsinformation verbessert wird. Schließlich werden europäische Sprachenvielfalt und kulturelle Repräsentation im digitalen Bereich gefördert, indem die europäische Webdateninfrastruktur insbesondere das europäische Web abdeckt – im Gegensatz zur angloamerikanischen Tendenz der heutigen (Web-)Datensammlungen. Darauf können digitale Produkte und Dienste aufgebaut werden, die europäische Sprachen und Kulturen ausgewogener berücksichtigen.

Umwelt- und Sicherheitsaspekte

Der Bau von riesigen neuen Datenzentren sorgt vielerorts für Unmut. Die europäische Webdateninfrastruktur ist hingegen so angelegt, dass sie dezentral – verteilt auf bestehende öffentliche und private Rechenzentren – aufgebaut werden kann. Gleichzeitig wird durch die gemeinsame Nutzung einer zentralen Webdateninfrastruktur sichergestellt, dass nicht jedes Unternehmen im Alleingang weite Teile des Webs crawlen und indexieren muss, was ein erhebliches Energieeinsparungspotenzial mit sich bringt. Dies trägt zur Erreichung der Klimaziele Europas bei und unterstützt die Entwicklung nachhaltiger digitaler Dienstleistungen und Wertschöpfung.

Die europäische Webdateninfrastruktur bietet zudem Werkzeuge zur Erkennung und Bekämpfung von Desinformationskampagnen im digitalen Raum. Insbesondere können Netzwerke aufgedeckt werden, die gezielt und koordiniert Desinformationen

verbreiten. Der Einsatz von KI-Systemen zur Verbreitung von Desinformation erschwert die Aufdeckung solcher Kampagnen mitunter. Dank eines systematischen Webdatenzugangs über eine Webdateninfrastruktur wäre die Erkennung von Desinformationskampagnen in viel größerem Maßstab möglich als bisher.

Wie bereits erwähnt, wäre aufgrund der zunehmenden Unabhängigkeit Europas im digitalen Raum zudem eine effektivere und konsequentere Durchsetzung von europäischer Digitalregulierung möglich und zu erwarten, z. B. im Daten- oder Urheberrechtsschutz sowie im Wettbewerbsrecht.

5 Aufbau und Betrieb einer europäischen Webdateninfrastruktur: organisatorische, finanzielle und regulatorische Aspekte

Organisatorische Rahmenbedingungen

Um die Webdateninfrastruktur in den EU-Mitgliedsstaaten bereitstellen und betreiben zu können, bietet sich ihre Integration in ein bestehendes Netzwerk von Rechenzentren an. Aus heutiger Sicht ist der beste Ort hierfür das bereits bestehende Ökosystem aus europäischen Hochleistungsrechenzentren (EuroHPC) mit angebundenen KI-Fabriken (AI Factories, Antennas). So kann die vorhandene Supercomputing-Infrastruktur (Netzwerke, Speicher und Rechenleistung) ressourcen- und kosteneffizient genutzt und sehr rasch EU-weit skaliert werden. Mehrere Hochleistungsrechenzentren, die Teil bestehender AI Factories sind, sind bereits Partner der europäischen Open Web Search Initiative und betreiben aktuell den Prototypen des Open Web Index, der weiterhin europäisch finanziert und skaliert werden sollte.

Aufbau und Betrieb der europäischen Webdateninfrastruktur benötigen ausreichende und langfristige finanzielle Mittel. Hierfür muss im EU-Haushalt eine entsprechende zweckgebundene Grundfinanzierung bereitgestellt werden. Gleichzeitig müssen kontinuierlich Forschung und Entwicklung im Bereich Big Data, Web- und Information-Retrieval-Technologien (Technologien zur Informationserschließung) gefördert werden, da sich diese Fachgebiete rasant weiterentwickeln. Nur so kann sichergestellt werden, dass die europäische Webdateninfrastruktur auf dem neuesten Stand und damit nützlich für Forschende, Industrie, Start-ups usw. bleibt. Nicht alle Mittel müssen zwangsläufig aus öffentlichen Quellen stammen, da Teile dieser Infrastruktur auch im Rahmen einer öffentlich-privaten Partnerschaft (Public-private-Partnership) aufgebaut und betrieben werden können.

Finanzierung – Bedarf und Potenziale

Der Betrieb einer voll ausgebauten europäischen Webdateninfrastruktur erfordert Entwicklungs-, Wartungs-, Betriebs-, Speicher-, Rechen- und Netzwerkkapazitäten. Die Kosten für den Aufbau und die Beschaffung dieser Kapazitäten sowie für die Entwicklung, Wartung und den Betrieb der Infrastruktur werden auf etwa 200 Millionen Euro pro Jahr geschätzt. Für den Fall, dass Speicher- und Rechenkapazitäten über bestehende öffentliche Rechenzentren bereitgestellt würden, minimierten sich diese Kosten auf etwa 50 Millionen Euro pro Jahr. Die ressourcen- und kosteneffizienteste Art, die Webdateninfrastruktur einzurichten, zu betreiben und zu warten, wäre daher die Implementierung im EuroHPC-Ökosystem.

Zwar ist die europäische Webdateninfrastruktur als öffentliche Infrastruktur gedacht, dennoch sind verschiedene Monetarisierungsmodelle zur Refinanzierung denkbar: Einerseits ein „Freemium"-Modell, bei dem grundlegende Funktionen kostenlos zur Verfügung gestellt werden, während umfangreichere Funktionen der Datennutzung im Premiumpaket angeboten werden. Alternativ könnten auch Abonnement- oder Pay-per-Use-Modelle etwa für den intensiven Schnittstellen-(API)-Zugriff zum Einsatz kommen, die je nach Nutzung flexibel abgerechnet werden.

Zusätzlich zum Kernangebot der Webdateninfrastruktur können Dienstleistungen als *value added services* angeboten werden. Dazu gehören zum Beispiel Wissensgrafen, Crawler-Dienste, spezialisierte Suchmaschinen, geografische Analysen, Business-Analytics-Dienste und das Training von Large Language Models (LLMs) mit Daten, die den europäischen Werten und Regulierungen entsprechen. Diese Mehrwertdienste helfen besonders dem digitalen Mittelstand sowie öffentlichen Projekten, niederschwellig und mit geringen Anfangsinvestitionen hochwertige Web- und Digitaldienstleistungen zu entwickeln und anzubieten.

Eine weitere mögliche Refinanzierungsquelle ist ein europäischer Datenmarktplatz. Anbieter*innen von Webinhalten und Urheber*innen können hierüber vergütet werden. Kommerzielle Nutzende (z. B. KI-Hersteller*innen, Betreiber*innen von Onlinemarktplätzen oder Suchmaschinen) können über den Marktplatz Mikrozahlungen zum Beispiel an Journalist*innen und Medienhäuser oder Kreativschaffende tätigen. Während der Hauptteil dieser Zahlungen an die Erstellenden von Webinhalten fließt, könnte ein Bruchteil der gezahlten Beiträge für die Vermittlung der Transaktionen und die dafür notwendige Infrastruktur eingesetzt werden.

Auf Basis der unterschiedlichen Maßnahmen für eine Refinanzierung (direkte Wertschöpfung) sowie der positiven wirtschaftlichen und gesellschaftlichen Auswirkungen der europäischen Webdateninfrastruktur (indirekte Wertschöpfung) prognostiziert eine Marktstudie bereits ab dem vierten Jahr nach Betriebsaufnahme eine positive Kapitalrendite.

Regulatorische Anpassungen

Zusätzlich zu den organisatorischen und finanziellen Voraussetzungen müssen die rechtlichen Grundlagen für den Betrieb einer europäischen Webdateninfrastruktur geschaffen werden. Die derzeitigen europäischen Rechtsrahmen bergen Risiken sowohl für den Betreiber einer europäischen Webdateninfrastruktur als auch für deren kommerzielle Nutzende. Zu klärende rechtliche Kernaspekte betreffen öffentlich zugängliche Webinhalte, die systematisch (in großem Umfang) über die europäische Webdateninfrastruktur heruntergeladen, verarbeitet und geteilt und später für kommerzielle Zwecke genutzt werden. Zu den gesammelten und geteilten Webinhalten gehören personenbezogene Daten, urheberrechtlich geschützte Inhalte und möglicherweise illegale Informationen. Der rechtskonforme Umgang mit solchen Daten muss daher klar geregelt werden.

Bei der Erhebung von Webdaten und deren Einbindung in ihre Dienste verstoßen große Technologieunternehmen aus Übersee regelmäßig gegen geltendes europäisches Recht und kalkulieren die mit möglichen Verstößen verbundenen Rechtskosten und Geldbußen ein. Der Mangel an Souveränität erschwert es der Europäischen Union, ihre Vorschriften im Bereich der Digitalisierung und des Datenschutzes wirksam durchzusetzen. Um gleiche Wettbewerbsbedingungen zu schaffen, muss Europa den rechtlichen Rahmen für den Aufbau und Betrieb einer öffentlichen Webdateninfrastruktur sowie deren kommerzielle Nutzung schaffen.

Regulatorische Klarstellungen sind vor allem in der Digital-Single-Market-Richtlinie, dem Digital Services Act (DSA) und der Datenschutz-Grundverordnung (DSGVO) erforderlich. Sie sollten möglichst eng gehalten werden, dabei aber den rechtssicheren Betrieb der Webdateninfrastruktur sicherstellen. Haftungsrisiken sollten für die öffentliche Webdateninfrastruktur als sogenannter „Intermediär“ von Webinformation ausgeschlossen werden (vorausgesetzt, ihre Betreiber handeln mit angemessener Sorgfalt und nicht vorsätzlich rechtswidrig). Zudem sollte eine rechtliche Grundlage für die umfassende Sammlung und Weitergabe öffentlich zugänglicher Webdaten durch die Webdateninfrastruktur geschaffen werden.

Eine wichtige Voraussetzung für die Anwendbarkeit der genannten regulatorischen Anpassungen ist die Verabschiedung von Standards und Prüfverfahren, an die sich die europäische Webdateninfrastruktur halten muss. Dazu gehören Transparenz und Berichterstattung über die gecrawlten bzw. gespeicherten Websites, klar definierte Verfahren zur Entfernung von Inhalten auf Antrag sowie systematische externe und

öffentliche Prüfungen, um Neutralität zu gewährleisten und Missbrauch zu verhindern. Die Gestaltung der europäischen Webdateninfrastruktur muss gewährleisten, dass die Grundrechte der Nutzenden geachtet werden und ihre Privatsphäre geschützt wird.

Darüber hinaus benötigen die Nutzenden der Webdateninfrastruktur Rechtssicherheit darin, wie sie innovative digitale Dienste und europäische Alternativen zu den dominierenden Anwendungen aus den USA auf rechtskonforme Weise aufbauen können. Dies erfordert eine solide Abwägung der Interessen von Anwendungsentwickler*innen und Webdienstanbieter*innen einerseits und den Urheber*innen von Inhalten sowie den betroffenen Personen andererseits.

6 Fazit und Handlungsempfehlungen

Der wirtschaftliche und gesellschaftliche Mehrwert einer europäischen Webdateninfrastruktur ist extrem hoch. Sie stärkt Deutschlands und Europas Digitalwirtschaft, trägt erheblich zur digitalen Souveränität bei und ermöglicht vielfältige Innovationen im kleinen wie im großen Maßstab. Sie ermöglicht die Entwicklung und den Betrieb zahlreicher deutscher und europäischer digitaler Dienste, die mit den Angeboten der großen US-amerikanischen Technologiekonzerne konkurrieren können.

Der Ausbau dieses Ökosystems rechtskonformer europäischer digitaler Dienste wird die Europäische Kommission und die Mitgliedsstaaten außerdem in die Lage versetzen, ihre digitalen Vorschriften gegenüber den nichteuropäischen Tech-Giganten besser durchzusetzen.

Deutschland muss sich dringend und verstärkt in Europa für eine solche öffentliche europäische Webdateninfrastruktur einsetzen. Erstens muss Deutschland helfen, die benötigten **rechtlichen Klarstellungen** zu erwirken, damit eine europäische Webdateninfrastruktur rechtssicher geschaffen und betrieben werden kann. Zweitens muss Deutschland dazu beitragen, die **langfristige Finanzierung** einer europäischen Webdateninfrastruktur mit zugehöriger Forschung auf europäischer Ebene sicherzustellen. Drittens muss Deutschland gemeinsam mit den weiteren europäischen Mitgliedsstaaten und Organen passende **Governance-Strukturen** für die europäische Webdateninfrastruktur entwickeln, aufbauen und langfristig betreiben.

Im Rahmen des OpenWebSearch.EU-Projekts wurden bereits die Grundlagen sowie ein Prototyp der europäischen Webdateninfrastruktur geschaffen. Nun muss diese Infrastruktur weiterentwickelt, operationalisiert und skaliert werden, um ein digital souveränes Deutschland und Europa zu schaffen.

Die Autor*innen

Ursula Gmelch hat Sprachwissenschaften sowie Interdisziplinäre Osteuropastudien an den Universitäten Augsburg und Regensburg studiert. Sie lebte in Kanada, Tschechien und Armenien. Bei der Open Search Foundation arbeitet sie seit September 2024 am Aufbau der Open Web Search Community.

Dr. Stefan Voigt studierte Geografie, Physik, Fernerkundung und Raumplanung in München und promovierte an der Universität Bern im Bereich der Geowissenschaften. Er arbeitet seit 2017 aktiv an der Vernetzung von Forschungs- und Rechenzentren sowie weiteren Akteur*innen aus Industrie, Politik und Zivilgesellschaft, um eine verteilte und offene Internetsuche für Europa zu konzipieren. 2018 gründete er hierfür die Open Search Foundation e. V. mit und ist seither ihr Vorstandsvorsitzender und Geschäftsführer. Seit 2000 ist er wissenschaftlicher Mitarbeiter im Erdbeobachtungszentrum des Deutschen Zentrums für Luft- und Raumfahrt (DLR). Aktuell leitet er dort das OpenSearch@DLR-Projekt für unbeeinflusstes Suchen und Auffinden wissenschaftlicher Informationen im Web.

Die Open Web Search Initiative, koordiniert durch den **Open Search Foundation e. V.,** ist eine europäische Bewegung von Menschen und Organisationen, die gemeinsam die Grundlagen für einen unabhängigen, freien und selbstbestimmten Zugang zu Informationen im Internet schaffen. Grundlegend ist die Überzeugung, dass mehr Unabhängigkeit von einzelnen außereuropäischen Technologiekonzernen – auch und insbesondere beim Zugang zu Informationen und Daten im Internet – großen Mehrwert für die europäische wirtschaftliche Wertschöpfung, die gesellschaftliche Stabilität und Demokratie bedeuten würde.

Grundidee der Open Web Search Initiative sind der Aufbau und Betrieb einer gemeinschaftlichen europäischen Webdateninfrastruktur, um die digitale Souveränität und Wettbewerbsfähigkeit im Web zu stärken. Gemeinsam mit europäischen Rechenzentren, Wissenschaftszentren, Industrie und NGOs arbeitet sie seit 2018 daran, Europa zu inspirieren, sich diese souveräne Webdateninfrastruktur für Analytics, Websuche und KI zu schaffen. Im OpenWebSearch.EU-Projekt, gefördert durch Horizon Europe, wurde ein erster Prototyp einer solchen Webdateninfrastruktur (inkl. eines Open Web Index) entwickelt.



Impressum

Herausgeberin: Heinrich-Böll-Stiftung e. V.,
Schumannstraße 8, 10117 Berlin

Fachkontakt: Franziska Busse,
Referentin Digitale Ordnungspolitik
E busse@boell.de

Erscheinungsort: www.boell.de

Erscheinungsdatum: November 2026

Lizenz: Creative Commons (CC BY-NC-ND 4.0)

https://creativecommons.org/licenses/by-nc-nd/4.0/deed.de

Gestaltung: Stockmar+Walter

Covermotiv: unsplash / BoliviaInteligente

Die vorliegende Publikation spiegelt nicht notwendigerweise die Meinung der Heinrich-Böll-Stiftung wider.

Die Publikationen der Heinrich-Böll-Stiftung dürfen nicht zu Wahlkampfzwecken verwendet werden.

Weitere Studien und Schriften zum Downloaden unter:
www.boell.de/publikationen