

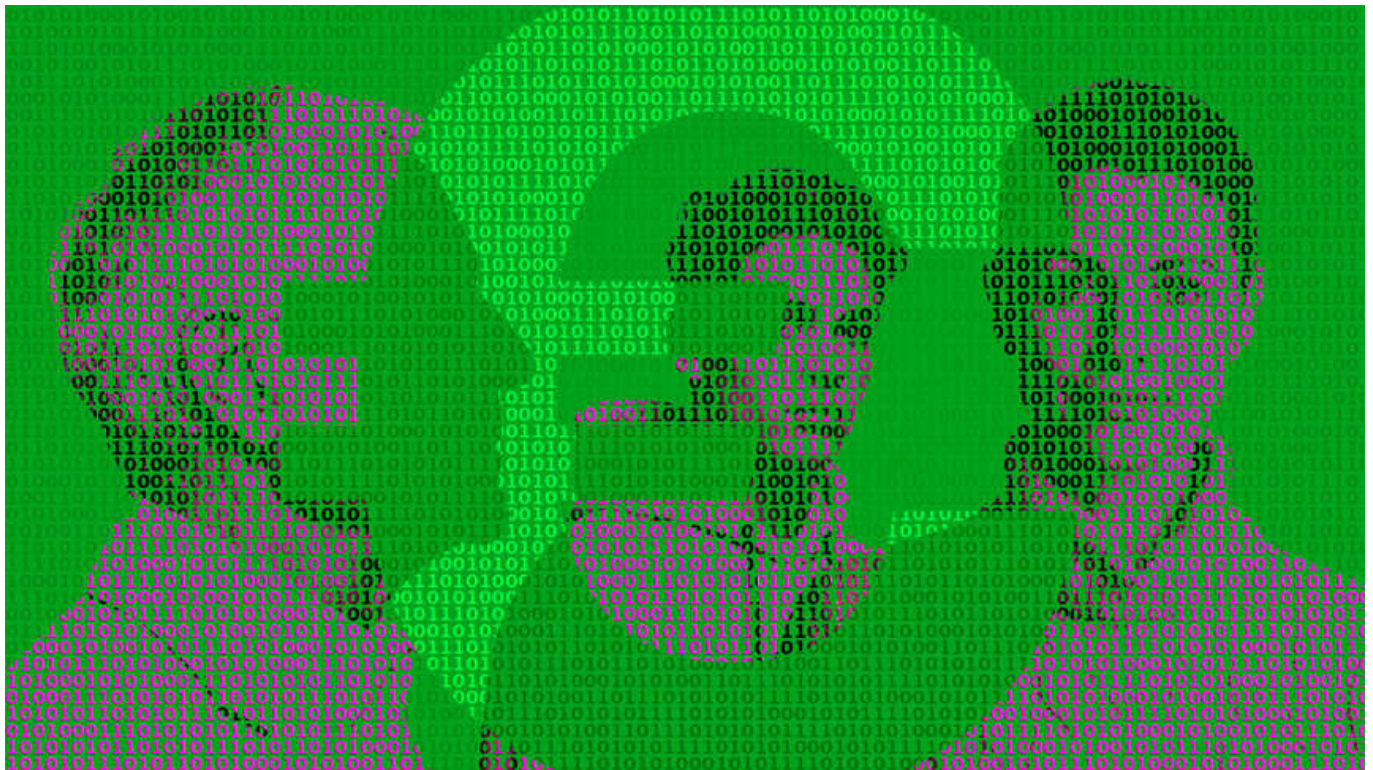


Bild: Ulrike Weis / KI / heise medien

Gesammelt, verknüpft, verkauft

Die unsichtbare Infrastruktur des Datenhandels

Data Broker machen Alltagsinformationen zur Ware. Der Handel mit Nutzerdaten setzt jährlich hunderte Milliarden Dollar um. Kaum einer kennt die Abläufe und die Akteure, obwohl so gut wie jeder mit seinen Online-Aktivitäten dieser Branche Rohstoff zuliefert.

Von Jo Bager

kompakt

- Klicks, Bewegungen, Einkäufe: Data Broker machen solche Alltagsinformationen zur handelbaren Ware.
- Die Branche macht Milliardenumsätze, riesige Datenpakete werden auf Onlinemarktplätzen verkauft.
- Staatliche Regulierung bekommt die Auswüchse des Handels nicht in den Griff.

Es ist 14:37 Uhr an einem Mittwochnachmittag. Eine Frau aus Hamburg öffnet auf ihrem Smartphone eine Seite auf einem Nachrichtenportal. Damit setzt sie eine globale Maschinerie in Gang. Die erfasst ihren Standort, ihr Gerät, ihre Browsereinstellungen und dutzende weitere Datenpunkte, analysiert und bewertet sie.

Eine Online-Auktionsbörse versteigert einen Werbeplatz auf der Nachrichtenseite, während sie lädt. Dabei annonciert sie die eingesammelten Daten in der Auktionsbeschreibung. Das meistbietende Unternehmen erhält den Zuschlag und zeigt ihr schließlich ein auf ihr Profil abgestimmtes Werbebanner an. Der gesamte Vorgang – vom Seitenaufruf über die Verbreitung der Daten an dutzende bietende Unternehmen bis zur Anzeige der personalisierten Werbung – dauert kürzer als ein Wimpernschlag und läuft völlig unbemerkt im Hintergrund ab.

Dieser Artikel beleuchtet die technische und ökonomische Infrastruktur, die diese Werbemaschinerie und viele weitere Kanäle mit Standortinformationen, Konsumgewohnheiten, Beziehungsstatus, Kreditwürdigkeit, Gesundheitsdaten, politischen Präferenzen und vielen weiteren Daten versorgt – sorgfältig gesammelt, verknüpft und in standardisierten Paketen zusammengeschnürt. Die Hauptakteure dieser Branche heißen Data Broker. Ihr Geschäftsmodell besteht darin, aus den digitalen Spuren, die Menschen im Alltag hinterlassen, handelbare Ware zu machen.

Milliardenmarkt im Dunkeln

Die ökonomischen Dimensionen dieser Industrie sind beeindruckend. Verschiedene Marktanalysen schätzen den globalen Data-Broker-Markt für das Jahr 2024 auf einen Wert zwischen 277 und 434 Milliarden US-Dollar. Das entspricht ungefähr dem Bruttoinlandsprodukt von Tschechien.

Wie diese Riesenumsätze zustande kommen, bleibt für die meisten Menschen allerdings im Dunkeln. Das liegt nicht nur an der technischen Komplexität der eingesetzten Systeme, sondern auch an der absichtlichen Intransparenz, mit der viele Akteure operieren. Schätzungen gehen davon aus, dass weltweit 4000 bis 5000 Data-Broker tätig sind. Die meisten dieser Unternehmen sind der Öffentlichkeit unbekannt.

Die Rohdaten, mit denen Data-Broker arbeiten, stammen aus einer Vielzahl von Quellen. Während öffentliche Register, Kreditauskünfte und Adressdatenbanken nach wie vor eine Rolle spielen, haben sich die Sammelquellen über die Jahre dramatisch erweitert. Heute werden Informationen überall abgeschöpft, wo digitale Geräte im Spiel sind, sei es das Smartphone, der PC, das Smart TV oder ein Wearable.

Datenspender Smartphone

Der wichtigste Erfüllungsgehilfe ist das Smartphone: Menschen haben es immer dabei und verbringen Stunden damit, auf Apps herumzuwischen. Moderne Smartphones sind vollgepackt mit Sensoren: GPS-Module, Beschleunigungssensoren, Gyroskope, Barometer, Helligkeitssensoren, Mikrofone, Kameras.

Jeder dieser Sensoren liefert Daten, die sich für Profiling-Zwecke auswerten lassen. Besonders wertvoll sind Standortdaten. Sie verraten nicht nur jederzeit, wo sich jemand aktuell aufhält, sondern erlauben Rückschlüsse auf Wohn- und Arbeitsort, Lebensstil, soziale Kontakte, Gesundheitszustand und politische Überzeugungen (siehe Artikel auf S. 60).

Dafür, dass Apps auf viele Daten zugreifen können, sorgen Analysedienste, Werbenetzwerke und Social-Media-Betreiber. Sie stellen Entwicklern Software Development Kits (SDKs) bereit. Das sind Programmbibliotheken, die App-Entwickler in ihre Anwendungen integrieren, um bestimmte Funktionen zu nutzen. Analysedienste zum Beispiel helfen, Schwachstellen in den Apps aufzuspüren; Werbung refinanziert den Programmieraufwand und den Betrieb der Apps. Die Apps übermitteln die erhobenen Daten nicht nur an ihre eigenen Entwickler, sondern auch an den SDK-Anbieter – und von dort aus häufig weiter an Data Broker.

Die meisten Nutzer wissen nicht, welche SDKs in ihren Apps integriert sind und welche Daten diese sammeln. Sie stellen höchstens fest, dass die Apps ungewöhnlich viele Rechte anfordern, Rechte, die sie eigentlich gar nicht benötigen – etwa wenn die Taschenlampen-App den Standort wissen und die Wetter-App das Telefonbuch auslesen will. Eine durchschnittliche Smartphone-App kann zehn oder mehr SDKs enthalten, die die Daten wild im Web herumfunken. Multipliziert mit dutzenden installierten Apps entsteht ein dichtes Netz von Datensammlern.

Welches Ausmaß der Datenabfluss über das Smartphone hat, zeigen die „Databroker Files“ – eine ausführliche Recherche von Netzpolitik.org gemeinsam mit dem Bayerischen Rundfunk und mehreren internationalen Medien. Die Kollegen haben mehrere Datensätze von Datenhändlern analysiert. Einer der untersuchten Datensätze enthält 380 Millionen Standortdaten aus 137 Ländern. Dabei handelte es sich nur um eine Kostprobe des Datenhändlers, die die Daten eines einzelnen Tages aus dem Juli 2024 umfasst.

Allein in Deutschland wurden an diesem Tag 795.000 Handys geortet. Insgesamt stammen die Daten aus rund 40.000 Apps. Die Android-App von Wetter Online bestimmte an dem Tag die Position zehntausender Nutzer „wohl teils auf den Meter genau“. Genaue Standortdaten gab es auch von Focus Online, Kleinanzeigen und FlightRadar24.

Sammelstelle Browser

Über das Tracking im Browser haben wir in der c't mehrfach ausführlich berichtet, zuletzt in [1], deshalb in aller Kürze: Die Werbebranche lässt sich immer neue Tricks einfallen, um Nutzer über einzelne Websites hinaus zu verfolgen.

Viele Jahre lang haben die Browser die Technik dazu selbst bereitgestellt: Third-Party-Cookies. Seit immer mehr Browser diese Cookies blockieren, kommen andere Techniken zum Einsatz. Dazu zählen verschiedene Fingerprinting-Verfahren, die versuchen, einen eindeutigen Fingerabdruck des jeweiligen Geräts zu erzeugen.

Fingerprinting war lange Zeit öffentlich geächtet und in Plattform-Policies verboten. Noch vor wenigen Jahren hatte Google Fingerprinting als Technik bezeichnet, die „die Wahlmöglichkeit der Nutzer unterläuft und falsch ist“. In der Praxis wurde Fingerprinting aber auch in der Vergangenheit weitgehend toleriert und heimlich genutzt. Seit Anfang 2025 erlaubt Google seinen Partnern nun offiziell digitales Fingerprinting.

Neben dem Fingerprinting nutzt die Werbeindustrie noch viele weitere Verfahren für das Tracking. Die Hersteller von Browsern und Werbeblockern versuchen ihrerseits, diese Maschen zu unterbinden: ein fortwährendes Katz-und-Maus-Spiel.

Quelle: Dr. Johnny Ryan

Verräterische Versteigerungen

Das sogenannte Real-Time Bidding (RTB) ist ein Auktionssystem für Werbeplätze. Wie im Beispiel am Anfang des Artikels werden dabei Werbeplätze auf Webseiten oder in Apps versteigert. Hinter RTB steckt ein präzise choreografierter Ablauf, der innerhalb von 100 Millisekunden abgeschlossen ist.

Wenn ein Nutzer eine Website aufruft, sammelt eine in die Website eingebettete JavaScript-Datei eines Werbenetzwerks clientseitige Daten: Gerätetyp, Betriebssystem, Bildschirmgröße, Geolokalisierung (GPS oder IP-basiert), Browser-Details sowie kontextuelle Signale wie die besuchte URL, Inhaltskategorie und aus dem Seitentext extrahierte Schlüsselwörter.

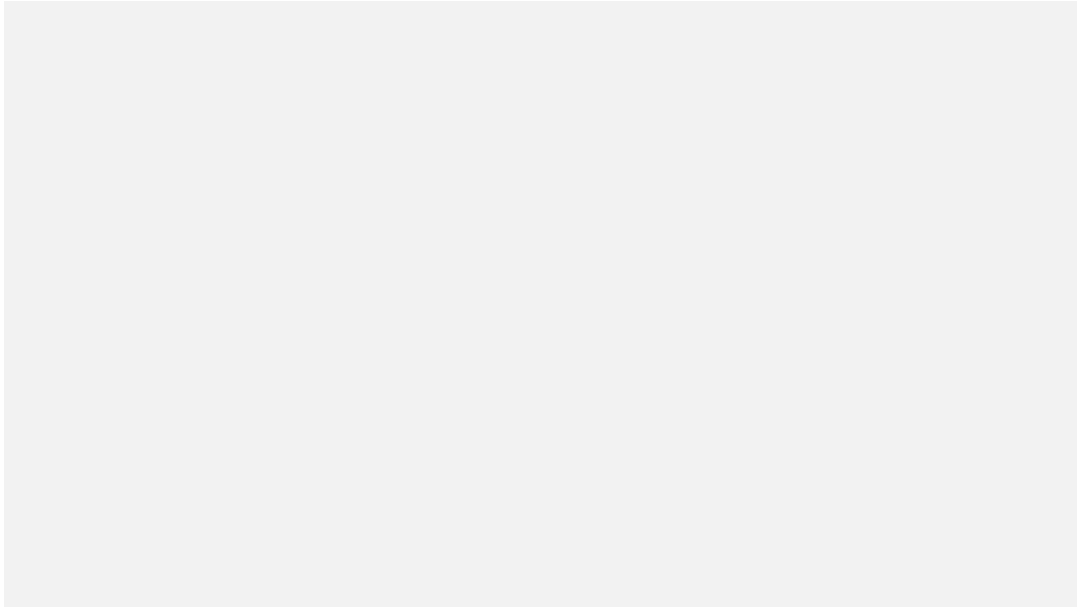
Diese Daten überträgt das JavaScript an eine sogenannte Supply Side Platform (SSP). Dabei handelt es sich um einen Dienstleister, der im Auftrag des Website-Betreibers Werbeflächen verwaltet und verkauft. Die Supply Side Platform (SSP) stellt eine Bid-Request zusammen, die alle gesammelten Nutzerdaten enthält. Diese Bid-Request sendet sie an die Auktionsplattform (Ad-Exchange), die sie an Dutzende Demand Side Platforms (DSPs) verteilt.

Die DSPs analysieren die Bid-Request daraufhin, welche Anzeige aus ihrem aktuellen Inventar am besten dazu passt und wieviel ihr ein Gebot für die Anzeige wert ist. Die Ad-Exchange empfängt Gebote von mehreren DSPs und ermittelt den Gewinner der Auktion. Der wird anschließend benachrichtigt und liefert die Anzeige an das Gerät des Nutzers aus.

Real-Time Bidding fungiert auch als massives Datenabflusssystem. Bei jeder Auktion werden Verbraucherdaten an potenziell dutzende Bieter gleichzeitig übertragen, obwohl nur ein Bieter – der Auktionsgewinner – diese Daten tatsächlich benötigt, um eine maßgeschneiderte Anzeige auszuliefern. Es gibt kaum technische Kontrollmechanismen, die sicherstellen, dass die anderen Parteien diese Daten nicht für unbeabsichtigte Zwecke behalten.

Die US-amerikanische Wettbewerbsbehörde FTC erhob 2024 Vorwürfe gegen den Data-Broker Mobilewalla, der sensible Standortinformationen von Verbrauchern oft ohne deren Zustimmung sammelte, behielt und diese Details mit Dritten teilte.

Die Beschwerde gegen Mobilewalla illustriert das Problem: Das Unternehmen soll Daten aus verlorenen Auktionen gesammelt und behalten haben, obwohl die Ad-Exchanges dies explizit verboten hatten. Da RTB-Auktionen Milliarden Mal täglich stattfinden und an jedem Vorgang dutzende Unternehmen beteiligt sind, entsteht ein unkontrollierbarer Datenstrom. Selbst wenn die Mehrheit der Beteiligten sich an die Regeln hält, reichen wenige schwarze Schafe, um riesige Datenmengen abzuschöpfen.



Datenraffinerien

Die Rohdaten aus Smartphones, Browsern und Auktionssystemen sind für sich genommen nur begrenzt wertvoll. Erst indem man sie verknüpft, anreichert und analysiert, entstehen die hochpreisigen Profile, die sich auf Datenmarktplätzen handeln lassen. Der ganze Prozess nennt sich Data Enrichment.

Dabei geht es darum, fragmentierte Datenpunkte aus verschiedenen Quellen einer einzigen Person oder einem Haushalt zuzuordnen. Eine Person hinterlässt Spuren auf verschiedenen Geräten (Smartphone, Laptop, Tablet), nutzt unterschiedliche Browser und Apps, wechselt zwischen WLAN und Mobilfunk, verwendet verschiedene E-Mail-Adressen und interagiert mit Dutzenden Diensten.

Die zentrale Herausforderung für Data Broker besteht darin, all diese Fragmente als zu einer Identität gehörig zu erkennen. Dabei spielen die bereits erwähnten Cookies, das Browser-Fingerprinting und ähnliche Techniken im Browser eine Rolle. Die Smartphones etwa stellen sogenannte Mobile Advertising IDs (MAIDs) bereit. Anhand seiner MAID lässt sich ein Gerät auch dann wiedererkennen, wenn man es mit verschiedenen Apps benutzt.

Darüber hinaus unterhält die Branche eine Reihe von Identifiern, um eine Person wiederzuerkennen. Dazu gehört die Unified ID 2.0. Anstelle von Cookies tritt eine pseudonyme Kennung. Sie basiert auf einer E-Mail-Adresse oder Telefonnummer des Nutzers, funktioniert also überall dort, wo sich der Nutzer anmeldet.

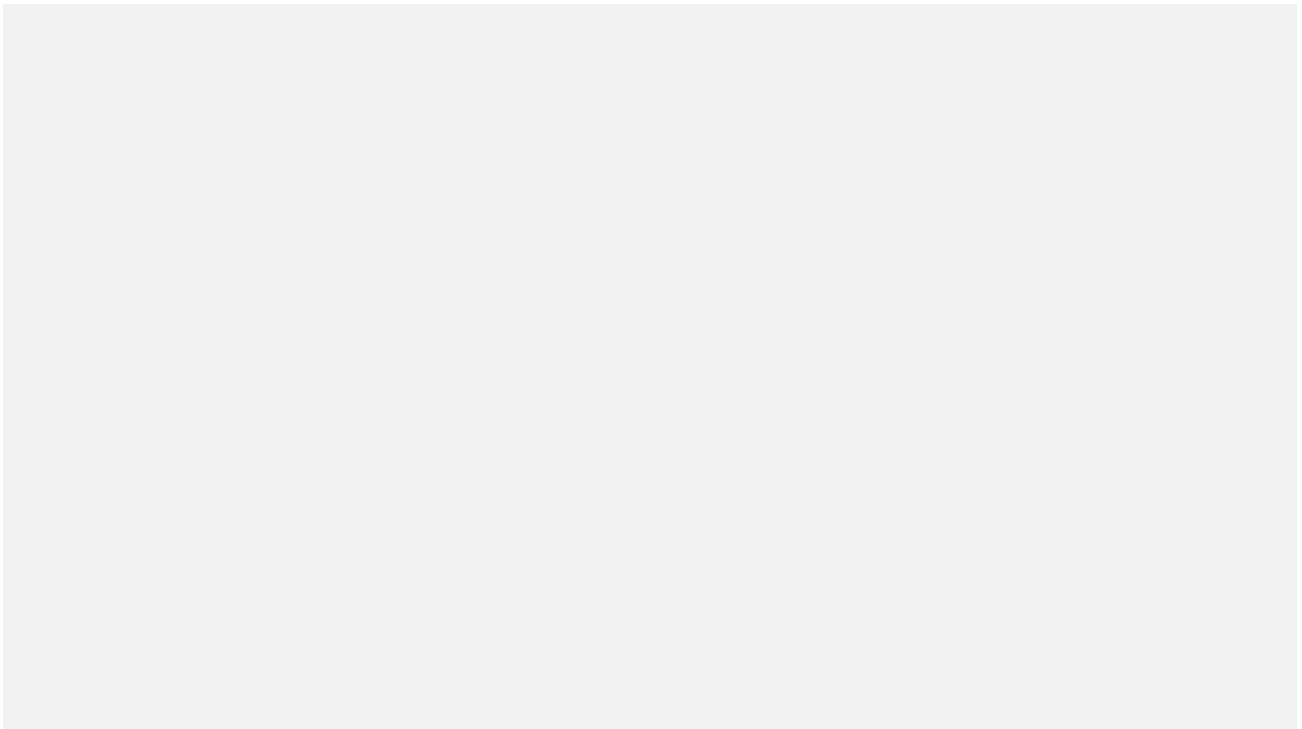
Stehen ihnen keine eindeutigen Identifier zur Verfügung, setzen Werbefirmen auf probabilistisches Tracking, also auf Wahrscheinlichkeitsberechnungen. Dabei sammeln sie Datenpunkte wie IP-Adressen, Browserversionen, Bildschirmauflösungen, Standortdaten, Zeitstempel sowie Nutzungsgewohnheiten und gleichen sie miteinander ab.

Wenn beispielsweise ein Smartphone und ein Laptop regelmäßig vom selben WLAN aus verwendet werden, ähnliche Websites zur gleichen Tageszeit aufrufen und vergleichbare Interessen zeigen, kann ein Algorithmus mit hoher Wahrscheinlichkeit schlussfolgern, dass beide Geräte derselben Person gehören. Dabei helfen auch Techniken wie „Continuity“ von Apple, „Chrome Tab Sync“ von Google, „Continue on PC“ von Microsoft, Samsungs „Flow“ et cetera. Solche Techniken wurden also nicht nur zur Bequemlichkeit der Nutzer entwickelt, sondern auch, um die Daten besser zusammenführen zu können.

Das Ergebnis sind sogenannte Unified Customer Profiles oder 360-Grad-Sichten: Datenbankeinträge, die alle verfügbaren Informationen über eine Person bündeln. Solche Profile können hunderte Attribute umfassen: von demografischen Basisdaten über Kaufhistorien und Website-Besuchen bis hin zu daraus abgeleiteten Eigenschaften wie Kreditwürdigkeit, politischen Präferenzen oder Gesundheitsrisiken.

Um diese Matching- und Anreicherungsprozesse zu bewältigen, hat sich ein eigener Markt für Data-Enrichment-Software entwickelt. Ein typischer Workflow sieht so aus: Ein Unternehmen verfügt über eine Liste von E-Mail-Adressen seiner Newsletter-Abonnenten. Über ein Enrichment-Tool sendet es diese Adressen an einen Data Broker.

Dieser gleicht sie mit seinen Datenbanken ab und liefert zusätzliche Informationen zurück – zum Beispiel Alter, Geschlecht, Wohnort, geschätztes Einkommen, Familienstand, Interessen sowie bisherige Käufe bei anderen Händlern. Das Unternehmen kann nun sein Marketing präziser ausrichten: individuelle Angebote für verschiedene Einkommensgruppen, unterschiedliche Botschaften für unterschiedliche Lebensphasen.



Die Daten-Supermärkte

Für den Vertrieb von Datenpaketen hat die Branche eine eigene Infrastruktur entwickelt: Auf großen Marktplätzen listen Anbieter wie bei Amazon ihre Datenpakete mit ausführlichen Beschreibungen. Käufer können das Angebot durchsuchen, filtern und Pakete direkt mit wenigen Klicks erwerben.

Einer der größten dieser Datenmarktplätze stammt aus Deutschland und nennt sich Datarade. Er zeigt stellvertretend, wie solche Marktplätze funktionieren. Mehr als 500 Anbieter stellen dort Datensätze in mehr als 500 Kategorien bereit, von „Geospatial Data“ über „Consumer Data“ und „Political Data“ bis hin zu „Location Data“.

Die Homepage ist so ähnlich aufgebaut wie ein klassischer Webshop für physische Waren. Der Interessierte kann das Angebot im Volltext durchsuchen oder sich Daten einer Kategorie auflisten lassen. Es gibt ausgefeilte Filteroptionen, in denen man die Suche zum Beispiel auf einzelne Anbieter einschränkt oder auf Datensätze, die bestimmte Länder umfassen.

Zu jedem Datensatz hält Datarade einen ausführlichen Steckbrief bereit, der die Größe, die erfassten Länder, die Kosten sowie ein Daten-Dictionary enthält. Letzteres schlüsselt detailliert die einzelnen Felder auf, gibt also einen Überblick über die ermittelten Merkmale und Attribute. Wer sich kostenlos auf der Plattform anmeldet, dem zeigt die Übersicht bei vielen Datensätzen einige echte Beispieldaten an. Manche Anbieter stellen auch größere kostenlose Beispielpakete bereit. Per Kontaktformular können Interessierte sich schnell mit den Händlern kurzschließen, wenn sie noch Fragen haben.

Datarade bietet Datensätze in verschiedenen Formaten an, die sich nach dem Verwendungszweck richten. Für Analysten und Datenspezialisten stehen strukturierte Formate wie CSV, JSON oder Excel-Dateien zum Download bereit. Diese eignen sich für statistische Auswertungen, etwa um Markttrends zu analysieren oder Kundenprofile zu erstellen. Wer größere Datenmengen verarbeiten möchte, kann auf Datenbank-Dumps oder spezialisierte Formate wie Parquet zugreifen, die sich effizient in Data Warehouses integrieren lassen.

Für zeitkritische Anwendungen reichen statische Datensätze nicht aus. Hier kommen APIs ins Spiel – Programmierschnittstellen, die Echtzeitabfragen ermöglichen. Das typische Szenario sieht so aus: Ein Online-Shop erhält eine Bestellung. Vor der finalen Bestätigung sendet das System eine API-Anfrage an einen Data Broker: „Was wissen wir über diese E-Mail-Adresse?“ Binnen Millisekunden kommt die Antwort zurück – mit Informationen über geschätztes Einkommen, bisheriges Kaufverhalten, Zahlungszuverlässigkeit und Betrugsrisiko. Basierend darauf kann der Shop entscheiden, welche Zahlungsmethoden er anbietet oder ob er eine Bonitätsprüfung verlangt.

Solche Echtzeitabfragen sind mittlerweile Standard in vielen Branchen. Versicherungen nutzen sie, um Prämien dynamisch anzupassen. Banken checken Kreditanträge in Sekunden statt Tagen. Marketingplattformen personalisieren Website-Inhalte basierend auf dem Profil des Besuchers. Die API-Infrastruktur macht Datenhandel unsichtbar und allgegenwärtig zugleich.

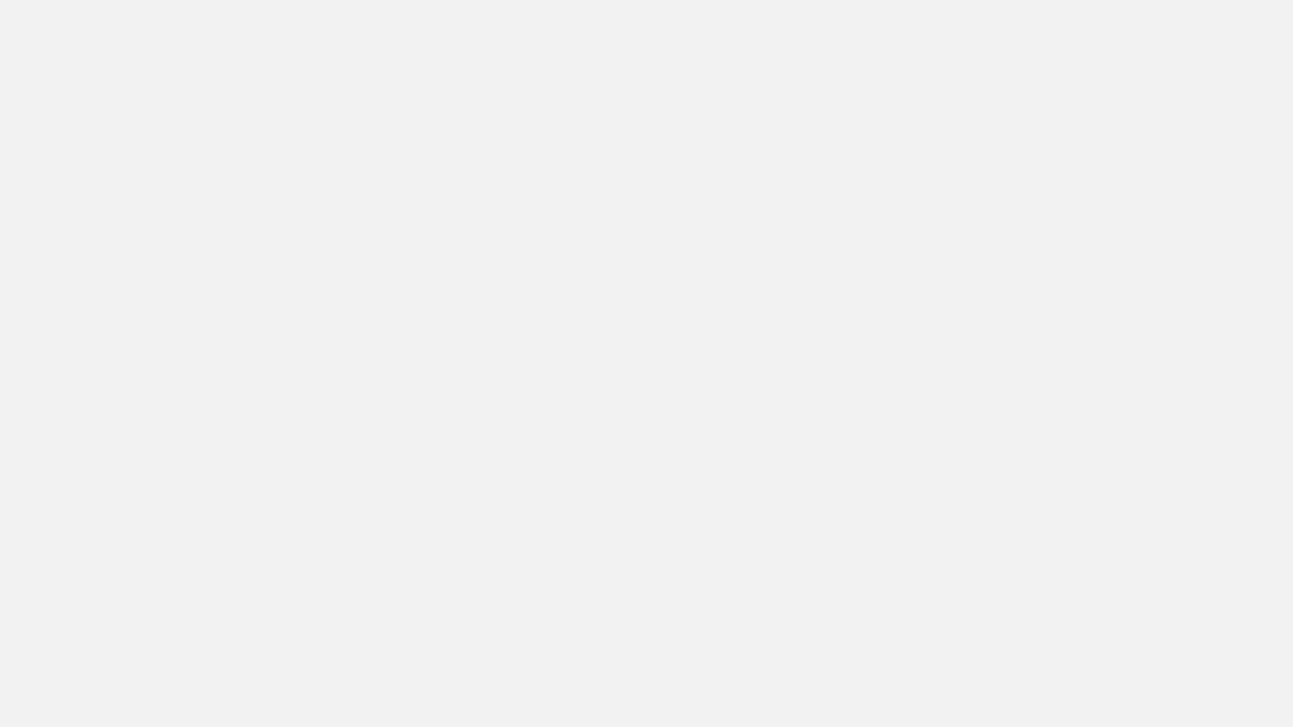
Die Preise und Preismodelle der bei Datarade gehandelten Datenpakete variieren stark. Manche Pakete kosten gerade einmal pauschal ein paar hundert Dollar. Andere werden nach der Anzahl der API-Calls oder Leads abgerechnet. Wenn also zum Beispiel ein Online-Shop mithilfe der Daten einen Verkauf generiert, sind ein paar Cent oder US-Dollar fällig. Der permanente Zugriff auf eine wirklich große Quelle mit den Daten von hunderten Millionen Menschen kann aber auch schon mal 10.000 US-Dollar oder mehr pro Monat kosten.

Welch brisante und weitreichende Informationen sich über solche Marktplätze beschaffen lassen und wie leicht dies geht, zeigte die „Databroker Files“-Recherche. Netzpolitik.org erhielt im Sommer 2024 3,6 Milliarden Standortdaten von Menschen in Deutschland – einfach so, als Gratis-Kostprobe. Die Daten enthielten Bewegungsprofile von rund 11 Millionen Geräten und ermöglichten es, Menschen bis in ihre Wohnungen zu verfolgen oder Besuche in Entzugskliniken, Bordellen oder bei Psychiatern nachzuvollziehen.

Besonders verhängnisvoll: Auch Bewegungsprofile von Mitarbeitern deutscher Geheimdienste und des Verteidigungsministeriums waren im Datensatz enthalten – ein erhebliches Sicherheitsrisiko für die nationale Sicherheit. Der US-amerikanische Datenanbieter Datastream Group hatte diese Geoinformationen für ein monatliches Abonnement von etwa 14.000 US-Dollar angeboten – stündlich aktualisiert, aus bis zu 163 Ländern. Der Fall macht deutlich: Was auf solchen Marktplätzen wie ein harmloses Geschäftsmodell wirkt, kann in Wirklichkeit demokratiegefährdende Dimensionen annehmen. Die Bundesregierung und Datenschützer forderten als Konsequenz ein grundsätzliches Verbot des kommerziellen Datenhandels mit Standortdaten.

Die Hersteller der Mobilfunkbetriebssysteme, Apple und Google, haben begonnen, auf das immer weiter ausufernde Datensammeln zu reagieren. Apple führte mit iOS 17 sogenannte Privacy Manifests ein – Deklarationen, die SDK-Entwickler zwingen, ihre Praktiken diesbezüglich offenzulegen. Google entwickelt mit der SDK Runtime für Android ein Sandboxing-System, das SDKs von Geräte-Identifikatoren abschirmt. Diese Maßnahmen zeigen, dass selbst die Plattformkonzerne das ungehemmte Abgreifen von Informationen auf ihren Plattformen als problematisch erkannt haben. Ob diese Maßnahmen wirklich Abhilfe schaffen, ist aber fraglich.

Auch von staatlicher Seite müssen die Datenhändler bisher kaum Beschränkungen fürchten. In dem Markt gibt es ein Flickwerk internationaler und US-bundesstaatlicher Gesetze mit vielen Schlupflöchern, das den Datenhandel nicht wirksam reguliert (siehe S. 72).



Jeder Datensatz wird fein säuberlich in einem Datenblatt beschrieben.

Fazit

Der Datenhandel ist eine weitverzweigte, hochprofitable Maschinerie, die aus digitalen Spuren Profile macht, aus Profilen Prognosen und aus Prognosen Profit. Mit tausenden Akteuren und einer ausgereiften technischen Infrastruktur hat sich eine Branche etabliert, die systematisch persönliche Informationen in handelbare Ware verwandelt. Staatliche Stellen scheitern bisher daran, diesen Markt wirksam zu regulieren.

Ganz entziehen kann man sich diesem Apparat als Einzelner nicht mehr – dafür ist er viel zu eng mit viel zu vielen Bereichen des Lebens verzahnt. Damit die Situation besser wird, müssen die Gesetzgeber die Datenhändler enger an die Leine nehmen. Wenn Sie selbst etwas beitragen und Druck auf die Politik ausüben

möchten, unterstützen Sie Organisationen wie Noyb (None of your business), EDRi (European digital rights) oder EFF (Electronic Frontier Foundation). Sie setzen sich dafür ein, dass der Datenhandel transparenter und besser von demokratisch legitimierten Organen kontrolliert wird. (jo@ct.de)

Databroker Files und weitere Infos: ct.de/y5mj

► Kommentieren



Leserbrief schreiben



Artikel als PDF herunterladen



Auf Facebook teilen



Auf X teilen

[Kontakt](#)

[Impressum](#)

[Datenschutzhinweis](#)

[Nutzungsbedingungen](#)

[Mediadaten](#)

[Verträge kündigen](#)



Heftinhalt

Lesezeichen



vorheriger Artikel

nächster Artikel

