

Niels Lepperhoff, Björn Petersdorf

Datenschutz bei Webstatistiken

Ergebnisse einer empirischen Analyse

Wer protokolliert das Surfverhalten im World Wide Web? Werden Besucher über eine Datenerhebung informiert? Wer kann technisch Bewegungsprofile mit Namen verknüpfen? Diesen Fragen gingen die Autoren des folgenden Beitrags in einer empirischen Untersuchung von über 14.000 Webpräsenzen von Unternehmen und Gemeinden nach.

Einleitung

Wer eine Webpräsenz betreibt, investiert (viel) Zeit und Geld. Unternehmen wollen wissen, ob dieses Geld wirklich produktiv und effizient investiert wird. Eine Erfolgskontrolle von Webseiten ist deshalb für einen wirtschaftlichen Betrieb unverzichtbar. Mit Hilfe von Webstatistiken – auch Web Tracking, Web Analytics oder Webcontrolling genannt – messen Unternehmen das Verhalten ihrer Besucher. Dienstleister erstellen Webstatistiken für Betreiber.

Webstatistiken geben aggregierte Informationen über die Besucher von Webseiten wieder. Sie beantworten u.a. folgende Fragen:

- Über welche Wege betreten Besucher die Webpräsenz?
- Wie viele Besucher hat die Webpräsenz?

- Was unternehmen Besucher auf der Webpräsenz?

Der Gesetzgeber trägt diesem Wunsch nach Erfolgskontrolle in § 15 TMG Rechnung. Nutzungsdaten, die Grundlage der Webstatistiken bilden, dürfen „für Zwecke der Werbung, der Marktforschung oder zur bedarfsgerechten Gestaltung der Telemedien“ (§ 15 Abs. 3 TMG) verwendet werden, wenn diese mit Pseudonymen arbeiten und der Besucher nicht widersprochen hat. Die Bewegungsprofile dürfen aber nicht mit personenbezogenen Angaben, die bspw. für die Nutzung eines Angebots, einer Bestellung oder eine Kontaktaufnahme erhoben werden, zusammengeführt werden (§ 15 Abs. 3 TMG).

Zwei Akteure können eine solche Zusammenführung vornehmen: der Webseitenbetreiber und sein Statistikersteller. Die erste Möglichkeit ist offensichtlich und soll hier nicht weiter betrachtet werden. Weil ein Statistikersteller die Webseiten verschiedener Betreiber betreut, bedroht eine Verknüpfung der Bewegungsprofile mit personenbezogenen Daten die Anonymität der Besucher deutlich stärker.

Betreiber regelmäßige statistische Auswertungen, die keinen Personenbezug mehr enthalten.

Werden nur Logfiles verwendet, bleiben Datenerhebung und Versand unter der Kontrolle des Betreibers. Anders sieht es bei einer eigenständigen Datenerhebung durch den Statistikersteller aus. Hier bindet der Betreiber in alle Webseiten Webpixel oder speziellen Script-Code ein, der die Daten für den Statistikersteller sammelt und direkt an diesen sendet. Meistens werden zusätzlich Cookies eingesetzt. Welche Daten gesammelt werden, entscheidet und kontrolliert der Statistikersteller. Deshalb hat der Betreiber keine Kontrolle über Datenerhebung, Speicherung, Auswertung und weitere Nutzung der Daten. Weil es sich rechtlich um eine Datenverarbeitung im Auftrag handelt (§ 11 BDSG), bleibt die Verantwortung beim Betreiber. Die Art der Verarbeitung ist deshalb vertraglich zu regeln.

Betreibt ein Statistikersteller Dienstleistungen, bei denen sich eine Vielzahl von Menschen persönlich registriert, erhält er zusätzlich personenbezogene Daten. Das Beispiel Google zeigt das Potential auf.

Google betreibt den Webstatistikdienst Google Analytics. Dieser Dienst setzt die Cookies unter der jeweiligen Domain der überwachten Webpräsenz. Deshalb ist zu vermuten, dass Besucher für jede Webpräsenz eine neue ID-Nummer erhalten.

Google kennt sowohl die IP-Nummer, den Aufrufzeitpunkt und Browser des Besuchers. Zwei unterschiedliche ID-Nummern von Cookies deuten auf den gleichen Besucher hin, wenn die IP-Nummer und der Browser bei beiden Besuchen identisch sind und die Besuche nicht lange auseinander liegen. Damit wird die Erstellung eines übergreifenden Bewegungsbildes



Dr. Niels Lepperhoff

Geschäftsführer der Xamit Bewertungsgesellschaft mbH
Schwerpunkte: Datenschutz und IT-Sicherheit
E-Mail: niels.lepperhoff@xamit.de



Björn Petersdorf

Geschäftsführer der Xamit Bewertungsgesellschaft mbH
Schwerpunkte: Web Tracking und IT-Controlling
E-Mail: bjoern.petersdorf@xamit.de

1 Personalisierte Bewegungsprofile

Betreiber erhalten durch die Logfiles ihrer Server und ggf. gesetzte Cookies Einblick in Surfgewohnheiten ihrer Besucher. Weil aussagekräftige Auswertungen Fachwissen voraussetzen, nutzen Betreiber externe Dienstleister – im folgenden Statistikersteller genannt.

Ein Statistikersteller holt entweder das Logfile vom Betreiber ab oder erhebt die Daten selber. In beiden Fällen erhält der

möglich. Die Anwendbarkeit ist lediglich eine Frage der verfügbaren Rechenleistung.

Dieses Szenario könnte bereits bald Vergangenheit sein, da Google offenbar einen neuen Cookie testet [Heise Online 2007]. Dieser neue Cookie enthält eine eindeutige ID-Nummer, die Webseiten-übergreifend gültig ist. Hintergrund ist, dass Werbetreibende präzisere Berichte erhalten sollen, wie Besucher auf Werbung reagieren. Besucher lassen sich damit noch einfacher beobachten.

Google bietet neben seiner gleichnamigen Suchmaschine zahlreiche Dienste, wie Google Apps, Google Mail und Google Adwords für Unternehmen und Privatpersonen an, die eine persönliche Anmeldung verlangen. Damit ist eine Zuordnung von Cookie und Besuchernamen für das Unternehmen technisch einfach zu realisieren.

Die Autoren behaupten hier nicht, dass Google derzeit eine solche Zusammenführung vornimmt. Das Unternehmen ist jedoch aufgrund der erhobenen Daten dazu in der Lage und will seine Position im Werbemarkt weiter ausbauen. Haupteinahmequelle von Google ist zielgerichtete Werbung, die auf das Verhalten von Besuchern zugeschnitten ist.

Wie gravierend sich eine solche Zusammenführung auswirkt, hängt davon ab,

- welche Webseiten überwacht werden. Betrachtet ein Besucher bspw. Informationsseiten einer Krankenkasse über Krankheiten, liegt die Vermutung nahe, dass er (oder nahe Angehörige) an den auf den betrachteten Webseiten dargestellten Krankheiten leidet. Recherchiert ein Unternehmen auf (universitären) Webseiten nach bestimmten Forschungsergebnissen und Veröffentlichungen, liegt wiederum die Vermutung nahe, dass es an einem ähnlichen Thema arbeitet.
- wie viele Webseitenbetreiber den Statistikdienst nutzen.

2 Erhebungsmethodik

Wie häufig protokollieren Webseitenbetreiber das Surfverhalten? Welche Dienstleister nutzen sie? Werden die Besucher über die Protokollierung unterrichtet? Um diese Fragen zu beantworten, untersuchte Xamit zwischen August und September 2007 mehr als 655.500 Webseiten von 14.016 deutschen Unternehmen und Ge-

meinden. Für jede Webpräsenz wurde untersucht,

- welcher von 19 Statistikanbietern mit eigener Datenerhebung eingesetzt wird und
- ob die Besucher über die Datenerhebung informiert werden.

Neben 1.249 Gemeinden wurden dabei mittelständische Unternehmen aus unterschiedlichen Branchen berücksichtigt wie z.B.

- Verarbeitendes Gewerbe
- Handel, Instandhaltung und Reparatur von Kfz und Gebrauchsgütern
- Gastgewerbe
- Hotel mit Restaurant
- Grundstücks- und Wohnungswesen
- Gesundheitswesen
- Rechtsanwälte & Steuerberater
- Werbung
- Informationstechnik
- Unternehmensberatung

Jede Branche war mit 850 bis 2.050 Webpräsenzen vertreten. Analysiert wurden jeweils die bis zu 6.000 zuerst gefundenen Webseiten pro Webpräsenz.

Der Fokus der Untersuchung lag auf Webpräsenzen. Webshops blieben unberücksichtigt, weil sie teilweise ein in das Shopsystem integriertes Tracking-Verfahren verwenden.

Jeder Statistikersteller bindet eine charakteristische Zeichenfolge in die überwachten Webseiten ein, um den Seitenaufruf protokollieren zu können. Diese Zeichenfolge ist auf allen überwachten Webseiten identisch. Kommt eine solche Zeichenfolge auf einer Webseite vor, wurde dies als Überwachung durch den zugehörigen Statistikersteller gewertet.

Eine Belehrung über eine Protokollierung des Surfverhaltens enthält ebenfalls charakteristische Worte (Cookies, Browser usw.), um aussagekräftig zu sein. Nach diesen Worten wurde gesucht, um zu bestimmen, welche Webseiten informieren und welche nicht. Die Reihenfolge der Worte ist nicht relevant.

Google verlangt in § 8.1 seiner Nutzungsbedingungen [Google 2007], die Nutzung von Google Analytics an „prominenter“ Stelle zu dokumentieren. Google schreibt den Wortlaut dieser Information vor und behält sich ein Kontrollrecht vor. Ob die von Google vertraglich vorgeschriebenen Worte auf einer Webpräsenz, die Google Analytics nutzt, vorkommen, wurde analog untersucht.

Aus methodischen Gründen wurden nur Statistikersteller berücksichtigt, die

eine eigene Datenerhebung durchführen. Logfile-Analysen blieben deshalb außen vor.

Durch die maschinellen Analysen sind Fehlzuordnungen nicht auszuschließen. Stichprobenhafte Kontrollen zeigten allerdings keine gravierenden Fehler. Daher können die Ergebnisse als valide betrachtet werden.

3 Marktführer: Google

- ▶ 7% der untersuchten Webpräsenzen nutzen Google Analytics für die Webstatistik, 1% einen anderen Anbieter. Google ist mit seinem Produkt Google Analytics damit Marktführer in Deutschland.

Abbildung 1 zeigt deutlich, dass die Nutzungsintensität von Webstatistiken branchenabhängig ist. Spitzenreiter sind Werbeunternehmen mit einem Anteil von 17% für Google Analytics und 2% für andere Anbieter. Schlusslichter sind das Gastgewerbe, Gesundheitswesen und Rechtsanwälte und Steuerberater.

Angesichts der skizzierten Möglichkeit, Bewegungsprofile mit dem Namen zu verknüpfen, erscheint die Nutzung von Google Analytics in Branchen mit besonders sensiblen Informationen hoch problematisch. Das Gesundheitswesen ist eine solche Branche. Bereits die Information, dass Herr Max Mustermann sich eingehend mit den Öffnungszeiten eines Onkologen beschäftigt, reicht für die Vermutung aus, dass er an Krebs erkrankt ist. Derart sensible personenbezogene Daten stellt § 3 Abs. 9 BDSG unter einem besonders strengen Schutz.

4 Informationspflicht ignoriert

- ▶ In der Praxis ignorieren 99% der untersuchten Betreiber, die Google Analytics einsetzen, die Kennzeichnungspflicht gemäß § 8.1 der Nutzungsbedingungen für Google Analytics (Abbildung 2).

Laut DENIC waren am 2. Oktober 2007 unter der Toplevel-Domäne „.de“ mehr als 11 Mio. Domains registriert. Bei einem Marktanteil von 7% nutzten rechnerisch 770.000 Webpräsenzen Google Analytics, sofern man Mehrfachregistrierungen einer Webpräsenz unter verschiedenen Domains unberücksichtigt lässt. Danach ver-

heimlichen 762.000 Webpräsenzen die Überwachung durch Google Analytics.

Vor dem Hintergrund, dass bspw. Webseiten aus dem Gesundheitswesen Rückschlüsse auf den Gesundheitszustand eines Besuchers erlauben, erscheint die mangelhafte Unterrichtung der Besucher kein Kavaliersdelikt zu sein.

5 Datenschutzerklärung

► Eine Datenschutzerklärung, die sich mit der Protokollierung des Surfverhaltens beschäftigt, haben wir auf 7% der Webpräsenzen gefunden.

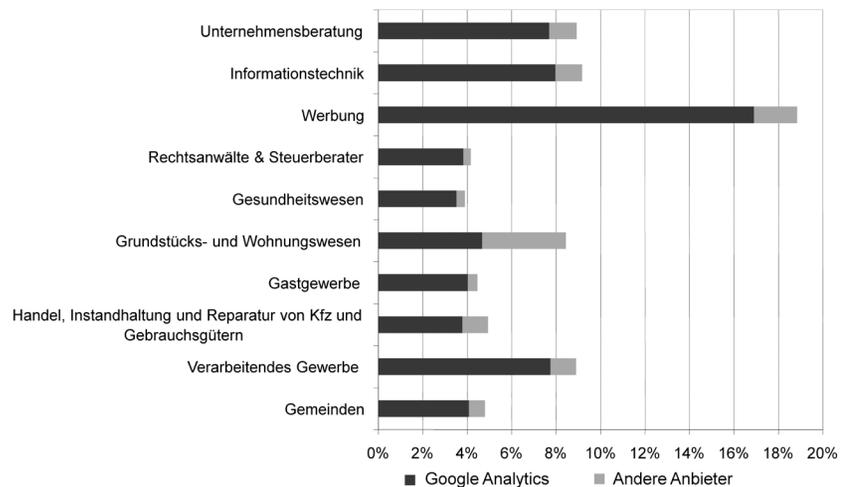
Es wird nicht unterschieden, ob die Datenschutzerklärung eine Protokollierung ausschließt oder auf sie hinweist. Die Information, dass Google Analytics oder ein ähnlicher Dienst genutzt wird, werten wir nicht als Datenschutzerklärung. Eine Datenschutzerklärung sollte über die Mitteilung, dass ein bestimmter Dienstleister beauftragt ist, hinausgehen.

Wieder zeigen sich deutliche Unterschiede zwischen den Branchen. Mit 13% führen Gemeinden, gefolgt von der Werbebranche mit 11%, das Feld an. Schlusslicht ist wieder das Gastgewerbe (Abbildung 3).

Unzureichende juristische Kenntnisse erklären die Unterschiede zwischen den Branchen nicht, denn nur 4% der Rechtsanwälte verfügen über eine Datenschutzerklärung. Vielmehr scheinen die Branchen unterschiedlich sensibel für den Datenschutz zu sein. Gemeinden verfügen i.d.R. über einen Datenschutzbeauftragten. In der Werbebranche entscheiden verschiedene Datenschutzregeln über die Zulässigkeit von Werbeaktionen. Entsprechend bekannter ist der Datenschutz. Der relativ hohe Anteil an Datenschutzerklärungen in der Informationstechnik stützt diese Annahme (9%), weil IT-Unternehmen sehr häufig mit Datenschutzfragen konfrontiert sind.

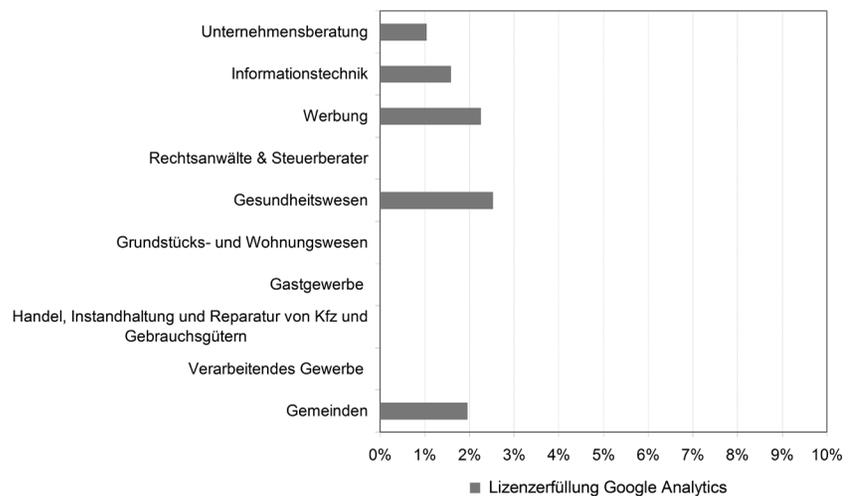
Akzeptieren wir einen Moment die Existenz einer Datenschutzerklärung als Indikator für Datenschutzsensibilität. Das Gastgewerbe, zu dem neben Restaurants und Gaststätten auch Hotels gehören, arbeitet täglich mit personenbezogenen Daten u.a. in Form von Reservierungen und Kreditkartenzahlungen. Liegt hier nur ein mangelndes Bewusstsein vor, oder auch ein mangelhafter Schutz personenbezogener Daten?

Abbildung 1 | Google Analytics ist Marktführer bei Webstatistiken mit eigener Datenerhebung



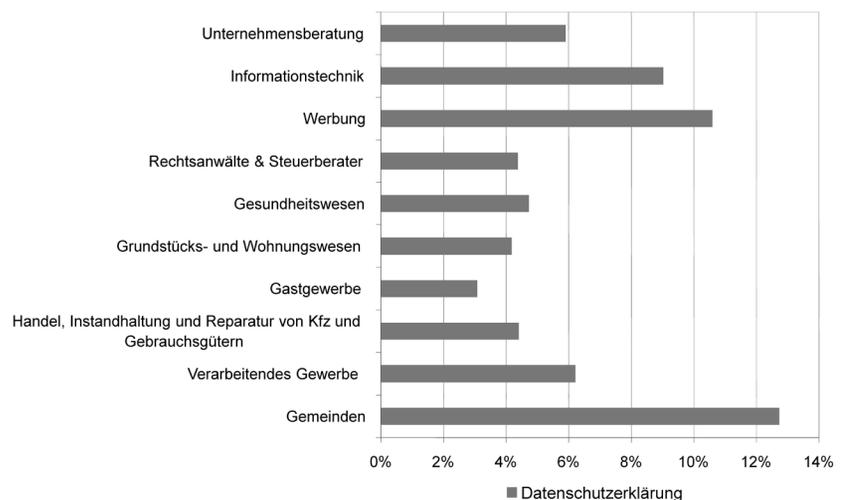
© Xamit Bewertungsgesellschaft mbH 2007

Abbildung 2 | Einhaltung der Kennzeichnungspflicht für Google Analytics



© Xamit Bewertungsgesellschaft mbH 2007

Abbildung 3 | Webpräsenzen mit Datenschutzerklärung



© Xamit Bewertungsgesellschaft mbH 2007

Es stimmt nicht zuversichtlich, dass auch Branchen wie das Gesundheitswesen (5%), Rechtsanwälte (4%) und Steuerberater (5%) nur unwesentlich häufiger eine Datenschutzerklärung publizieren. Zusammengefasst scheinen die meisten Unternehmen Datenschutz nicht ernst zu nehmen.

6 Selbstschutz

Unternehmen, die ein kundenfreundliches Image bevorzugen, sollten genau prüfen, welches Erhebungsverfahren für Ihre Webseitenbesucher sie einsetzen. Klare vertragliche Regelungen mit dem Statistikersteller sollten regeln, welche Daten erhoben werden, wie diese gespeichert und genutzt werden und dass ein Personenbezug technisch ausgeschlossen wird. Der Betreiber muss die Einhaltung auch kontrollieren und sollte einen verständlichen Hinweis auf seiner Webpräsenz nicht vergessen.

Wer einen externen Dienstleister beauftragt, sollte genau überlegen, welche Informationen dieser Dienstleister aus der Webstatistik entnehmen kann. Bspw. gibt

die Webstatistik eines Webshops Auskunft über den erzielten Umsatz. Auch sollten die Belange der Besucher berücksichtigt werden. Von einem Arzt erwartet ein Besucher unbedingte Geheimhaltung. Hier sorgt eine (intern) durchgeführte Logfile-Analyse für mehr Diskretion.

Bereits mit einfachen und kostenlosen Mitteln können Privatpersonen und Unternehmen ihre Surfspuren verringern:

- Browser so einstellen, dass Cookies höchstens für die aktuelle Sitzung angenommen werden
- Bei sensiblen Themen einen Anonymisierungsdienst verwenden
- Bei Nutzung von Firefox Scripte selektiv mit der Firefox-Erweiterung „noscript“ steuern, so dass Cookies von Google und Co. gar nicht erst gesetzt werden können. Ein vergleichbares Werkzeug ist uns für den Internet Explorer allerdings nicht bekannt.
- Keine Toolbar von Google, Yahoo, Alexis u.a. im Browser einsetzen, da die Toolbars das Surfverhalten protokollieren.

Fazit

Ein wirtschaftlicher Betrieb einer Webpräsenz setzt eine Erfolgskontrolle mittels Webstatistik voraus. Auf der anderen Seite wollen Besucher unbeobachtet surfen.

In diesem Spannungsfeld kommt der Auswahl eines geeigneten Dienstleisters für Webstatistiken eine zentrale Rolle zu.

In der Praxis scheinen Datenschutzüberlegungen bei der Auswahl eines Statistikdienstleisters keine oder nur eine untergeordnete Rolle zu spielen. Zudem erfahren Besucher meistens nicht, ob und wer ihr Surfverhalten protokolliert. Besucher können deshalb Unternehmen mit Datensammelwut weder erkennen noch diese meiden.

Literatur

- Google (2007): Google Analytics Bedingungen. URL: <http://www.google.com/analytics/de-DE/tos.html>. Stand: 2007-10-01.
- Heise Online (2007): Google testet neuen Cookie. 13.08.2007. URL: <http://www.heise.de/newsticker/meldung/94278>