International Working Group
on Data Protection
in Telecommunications

675.46.35

**Arbeitspapier und Empfehlungen
zu der Veröffentlichung personenbezogener Daten im Web,
der Indexierung des Inhalts von Websites und dem Schutz der Privatsphäre**

53. Sitzung, 15. – 16. April 2013, Prag (Tschechien)

- Übersetzung -

1. Hintergrund

Einer der wesentlichen Stützpfeiler des Datenschutzes war schon immer das Recht des Betroffenen, über seine Daten zu bestimmen. Ein wesentliches Element dieser Kontrolle ist das Recht, die eigenen Daten gelöscht zu bekommen, wenn sie rechtswidrig verarbeitet werden oder wenn der Betroffene ihrer Verarbeitung nicht länger zustimmt. Der kürzliche Vorschlag der Europäischen Kommission für einen neuen Regulierungsrahmen versucht, dieses Recht zu stärken, indem er ein „Recht auf Vergessen“ durch andere und im Web vorsieht. Dies gilt unbeschadet von solchen Fällen, in denen es ein legitimes und rechtlich gerechtfertigtes Interesse gibt, Daten veröffentlicht und sichtbar zu halten, wie etwa in Medienarchiven oder zum Zwecke historischer Aufzeichnungen, und es ist klar, dass das Recht auf Vergessen nicht a priori Vorrang vor dem Recht auf freie Meinungsäußerung oder der Medienfreiheit haben kann[1].

Angesichts der Struktur des Webs sind viele Einzelfragen im Hinblick darauf, wie ein solches „Recht auf Vergessen“ implementiert werden könnte, sowohl auf der technischen als auch auf der juristischen Seite immer noch ungelöst. Personenbezogene Daten (und jegliche andere Informationen), werden sehr wahrscheinlich öffentlich zugänglich bleiben, wenn sie einmal online veröffentlicht sind,. Sogar wenn sie auf der ursprünglichen Webseite gelöscht werden, können sie vor der Löschung auf anderen Seiten verlinkt oder gespiegelt werden. Das Web weiß nicht zu „vergessen“ und gegenwärtig ist kein einfaches technisches Werkzeug verfügbar, das die systematische Löschung von Daten im Web sicherstellen könnte (d. h., dem Web das Vergessen beibringen könnte). Kurz gesagt, es gibt keinen „Löschknopf“ und es ist zweifelhaft, ob es ihn jemals geben wird.

Dennoch gibt es bereits heute Wege, das Recht des Einzelnen auf Vergessen in einem gewissen Ausmaß zu schützen, indem man sich Werkzeuge zu Nutze macht, die Administratoren von Websei-

[1] The EU Data Protection Reform 2012: Making Europe the Standard Setter for Modern Data Protection Rules in the Digital Age, Viviane Reding SPEECH/12/26, Innovation Conference Digital, Life, Design, München, 22. Januar 2012, http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/12/26; für eine Kritik dieses Ansatzes s. Rosen, The Right to Be Forgotten, 64 Stan. L. Rev. Online 88

Secretariat
Berliner Beauftragter für
Datenschutz und Informationsfreiheit
An der Urania 4- 10
D-10787 Berlin
Phone +49 / 30 / 13889 0
Fax: +49 / 30 / 215 5050

E-Mail:
IWGDPT@datenschutz-berlin.de

Internet:
http://www.berlin-privacy-group.org

ten zur Verfügung stehen[2], um die freie Verfügbarkeit personenbezogener Daten zu begrenzen, wie auch durch Nutzbarmachung der Möglichkeiten von Suchmaschinen. Im gegenwärtigen Web könnte das Recht auf Vergessen[3] besser als ein „Recht, nicht gefunden zu werden" interpretiert und umgesetzt werden.

2. Die Aussichtender Nutzer, die Kontrolle über ihre personenbezogenen Daten im Web zurückzugewinnen

Die zunehmende Veröffentlichung personenbezogener Daten im Web in den letzten Jahren hat zu neue Herausforderungen und Risiken des Schutzes der Privatsphäre der Bürger Anlass hervorgerufen und gleichzeitig zur Verschärfung existierender Risiken geführt. Das Aufkommen sozialer Netzwerke hat in diesem Zusammenhang eine besonders wichtige Rolle gespielt[4].

Während in diesem Zusammenhang Technologien zur Förderung der Veröffentlichung und verfügbar machen von Daten – einschließlich personenbezogener Daten – im Web dramatische Fortschritte gemacht haben, scheint die Entwicklung von Technologien zur Kontrolle der Verfügbarkeit solcher Daten im Web immer noch in den Kinderschuhen zu stecken. Während Arbeiten für ein „policy-aware Web"[5] in der vergangenen Dekade stattgefunden haben, scheinen wir immer noch weit von jeglichen effektiven, einfach zu nutzenden und breit verfügbaren Werkzeugen entfernt zu sein, die es Bürgern ermöglichen würden, die Kontrolle über ihre eigenen Daten auch nur in einem begrenzten Maß (zurück-) zu gewinnen, wenn diese einmal im Web veröffentlicht worden sind.

Ein mögliches Entwicklungsziel für solche Technologien könnte die Förderung der Löschung aller Kopien von Daten auf jeglichen Geräten oder in jeglichen Speichern sein, in denen sie aufbewahrt werden. Gegenwärtig könnte dies wohl Probleme hinsichtlich der Skalierbarkeit aufwerfen (sogar wenn ein automatisierter Ansatz gewählt wird), besonders, wenn Daten im Laufe der Zeit von der Gemeinschaft der Nutzer im Web verbreitet, weiter verfeinert oder re-kontextualisiert worden sind. Es gibt gegenwärtig keine technische Möglichkeit, alle Kopien eines Objekts und Kopien von Informationen, die mit diesem Objekt im Web zusammenhängen, zu identifizieren und zu lokalisieren. Allerdings könnte dies in einem zukünftigen „policy-aware Web" möglich sein.

Für neu erzeugte Daten könnte die Verfügbarkeit im Web durch das Setzen von zeitlichen Begrenzungen (Verfallsdaten) im Bezug auf das jeweilige Objekt begrenzt werden. Dies kann auf vielen Wegen erreicht werden. Beispielsweise könnte man Daten mit „aktiver" (ausführbarer) Software verbinden, die interveniert, wenn das Verfallsdatum erreicht ist, um die Anzeige der Daten auf einem Bildschirm zu deaktivieren oder die Möglichkeit, Screenshots von einem Bild zu erstellen, zu blockieren oder den ursprünglichen Inhalt zu löschen oder zu verschlüsseln. Alternativ können Daten auch

---

[2] Eine solche Sammlung von Werkzeugen sind die Google Webmaster Tools, die es Webmastern ermöglichen, zu sehen, wie Google ihre Site durchsucht und indexiert, und es Webmastern ermöglicht, zu beeinflussen, wie die indexierten URLs angezeigt werden. Ein Link zu den Werkzeugen ist unter http://www.google.ca/webmasters/ verfügbar.

[3] Man beachte, dass der Ausdruck „Recht auf Vergessen" in diesem Papier in einem weiteren Sinne genutzt wird als in dem Entwurf der Datenschutzgrundverordnung der Europäischen Union, und dass dieses Papier keine Aussage enthält, ob ein „Recht auf Vergessen" in dieser Verordnung umgesetzt werden soll oder nicht.

[4] Vgl. Bericht und Empfehlung dieser Gruppe zum Datenschutz in sozialen Netzwerkdiensten – „Rom Memorandum" (Rom (Italien), 3. – 4. März 2008); http://www.datenschutz-berlin.de/attachments/470/675.36.13.pdf

[5] Für einige existierende Vorschläge zur Schaffung eines „policy-aware Web" vgl. Fußnote 27 auf Seite 10 des „Rome Memorandum" (Fußnote 4 oben). Das Konzept des policy-aware web kombiniert verschiedene existierende Technologien, namentlich strukturierte Daten, Identitätsmanagement, Zugriffskontrolle und „sticky policies" (d. h. Nutzungsregeln, die zusammen mit den Daten selbst verbreitet werden).

mit einem Verfallsdatum „markiert“ werden, sodass alle Server, die mit dem Objekt umgehen, dieses Datum berücksichtigen und die Daten nach dem Verfallsdatum entfernen können.

Weitere interessante Beispiele, wie die Lebenszeit neu generierter Daten im Web angepasst werden kann, werden von einigen anderen neu entstehenden Anwendungen gegeben. Zum Beispiel können Nutzer ein sicheres Overlay-Netz benutzen, das die Sichtbarkeit von Inhalten, wie z. B. einer Nachricht oder eines Bildes durch Nutzung von Ende-zu-Ende-Sicherheit und Zugriffskontrollregeln auf eine Gruppe beschränkt, die zu dem selben Overlay-Netzwerk gehört. In wiederum anderen Anwendungen bleibt eine Textnachricht im Mobilfunk bis zu einem bestimmbaren Verfallsdatum zu einem Nutzer verfügbar. Schließlich können „Nutzer-zentrierte“ Lösungen genannt werden, bei denen der legitime Eigentümer eines Datums selektiv Zugriff darauf gewähren kann, indem er Links zu dem Ort veröffentlicht, wo die Daten in Wirklichkeit nur in einem spezifiziertem Zeitraum gespeichert sind.

Diese Beispiele können als Bausteine für ein zukünftiges „policy-aware Web“ dienen. Allerdings ist eine Menge gründlicher Forschung und Entwicklung nötig, um diese Elemente zu effektiven Werkzeugen für den besseren Schutz der Privatsphäre der Bürger weiterzuentwickeln. Die Arbeitsgruppe ruft die relevanten Akteure in diesem Feld (Industrie, Wissenschaft und Regierungen) dazu auf, ihre Anstrengungen weiter zu verstärken, um hier Fortschritte zu machen.

3. Beschränkung der Verfügbarkeit personenbezogener Daten im Web durch Kontrolle ihrer Indexierbarkeit durch Suchmaschinen

Ein weiter Baustein zur Beschränkung der Verfügbarkeit und ein Beitrag zur Löschbarkeit von Daten im gegenwärtigen Web besteht in der Beschränkung ihrer Verfügbarkeit in den Ergebnissen von Anfragen bei Suchmaschinen[6]. Dies ist bereits jetzt technisch möglich und steht Website-Administratoren als Option zur Verfügung. Sie beruht im Wesentlichen auf zwei Alternativen: Dem „robots.txt-Protokoll“[7] und der Nutzung von an ein Objekt gebundenen Markierungen („tags“), um zu signalisieren, dass ein bestimmter Inhalt oder eine bestimmte Seite nicht von einer Suchmaschine indexiert werden soll.

Das „robots.txt-Protokoll“ arbeitet mit einem kleinen Satz von Instruktionen, die in einer Text-Datei codiert sind (der „robots.txt“-Datei), die im Wurzelverzeichnis einer Domain enthalten ist (z.B. http://example.com/robots.txt). Die Datei wird, falls sie vorhanden ist, von einem Crawler (einem Programm, das von Suchmaschinen genutzt wird, um eine Momentaufnahme einer Website zu erstellen) vor der Indexierung der jeweiligen Website gelesen. Die betreffenden Instruktionen erlauben es, *bestimmte Crawler* dazu aufzufordern, *bestimmte Dateien und/oder Verzeichnisse* auf der Website zu ignorieren. Die Instruktionen werden von Crawlern durch Textvergleich alphanumerischer Zeichenketten in der Reihenfolge ausgeführt, in der sie in der robots.txt-Datei enthalten sind. Zu den Anwendungsgrenzen des Protokolls zählen das Fehlen einer ausreichenden Skalierbarkeit, dass es mit ftp-Servern nicht funktioniert und dass die Information verloren geht, wenn Inhalte von einer Website kopiert werden[8].

Alternativ können verschiedene Kategorien von Markierungen („tags“) als Attribute einer spezifischen Web-Seite genutzt werden (aber auch in Verbindung mit individuellen Elementen einer spezi-

---

[6] S. auch Recommendation CM/Rec(2012)3 des Europarats zum Schutz der Menschenrechte in Bezug auf Suchmaschinen.

[7] Das „robots.txt-Protokoll“ wird auch als „Robots Exclusion Protocol“ und als „Robots Exclusion Standard“ bezeichnet. Das Protokoll ist in einem abgelaufenen „Internet Draft“ der IETF definiert, online verfügbar unter http://www.robotstxt.org/norobots-rfc.txt .

[8] Manchmal können außerdem Veränderungen bei Web-Inhalten und/oder Präferenzen bei der Indexierung nicht in Suchergebnissen widergespiegelt werden. Es hat sich als bedeutsames Problem erwiesen, Suchmaschinen dazu zu bringen, ihre Indizes zu aktualisieren.

fischen Seite, wie einem Bild oder einer Datei darin), um zu signalisieren, dass das Objekt/die Seite nicht in die Ergebnisse einer Suchanfrage aufgenommen werden sollte.

Es sollte betont werden, dass diese Ansätze beide vollständig auf Netz-Etikette (d. h. auf die Kooperation der betroffenen Parteien) basieren. Als solche sind sie nur sehr schwer durchzusetzen. Ihre Implementierung durch Websites und Einhaltung durch Suchmaschinen ist völlig freiwillig. Während sie die Risiken der Indexierung, die durch Verlinkung von Webseiten Dritter verursacht werden, abschwächen können, können sie nicht *per se* sicherstellen, dass ein bestimmtes Informationsobjekt niemals durch eine Suchmaschine indexiert werden wird, besonders, wenn dieses Objekt öffentlich zugänglich ist und von anderen Webseiten mit anderen Zugangsregeln für Crawler verarbeitet werden kann[9].

4. Empfehlungen für Website-Administratoren

Website-Administratoren spielen eine entscheidende Rolle in den beiden oben beschriebenen Kategorien der Löschung, und zwar durch ihre Möglichkeit, die Verfügbarkeit von Daten und die Indexierbarkeit von Objekten zu begrenzen. Um zu den o. g. Zielen beizutragen, gibt die Arbeitsgruppe die folgenden Empfehlungen:

- Betreiber von Websites sollten ihre Nutzer darüber informieren, welche personenbezogenen Daten sie aufbewahren und für welche Zwecke. Sie sollten ihren Nutzern einen einfachen Mechanismus für die Auskunft über ihre personenbezogenen Daten zur Verfügung stellen, und ihnen erlauben, diese zu berichtigen und/oder dauerhaft zu löschen, wie es in der existierenden Datenschutzgesetzgebung vorgesehen ist. Solche Auskunftsmechanismen sollten nutzerfreundlich sein und sollten nicht zu zusätzlichen Kosten für Nutzer führen oder ihnen ungerechtfertigte Verzögerungen oder praktische Belastungen aufbürden.

- Auf spezifische Anforderung eines Betroffenen, und wenn keine anderen legitimen Interessen oder gesetzlich bindende Beschränkungen existieren, sollten Webmaster die relevante Information umgehend von ihrer Website entfernen. Zusätzlich sollten sie Anbietern von Suchmaschinen signalisieren, den betreffenden Teil der Website zu re-indexieren, um die Daten auch aus dem Suchindex und existierende Kopien im Cache von Suchmaschinen löschen zu lassen.

- Webmaster sollten ihren Nutzern spezifische Werkzeuge zur Verfügung stellen, die es ihnen erlauben, ihre Indexierungs-Präferenzen für die Suche individuell anzupassen[10]. Alternativ könnte auch die Nutzung des „noindex“-meta-tag erwogen werden, dass in dem HTML-Code der betreffenden Seite oder in dem HTTP-Header eingebunden wird oder der sitemap.xml-Datei, um die relevanten Suchpräferenzen im Zusammenhang mit bestimmten Objekten zu signalisieren[11].

---

[9] S. in dieser Hinsicht auch die Empfehlungen, die im „gemeinsamen Standpunkt zu Datenschutz bei Suchmaschinen im Internet“, wie 1998 verabschiedet und 2006 überarbeitet, enthalten sind; http://www.datenschutz-berlin.de/attachments/237/WP_Suchmaschinen_de.pdf

[10] Vgl. den von der „blogger.com“-Plattform zur Verfügung gestellten Mechanismus, der es Nutzern ermöglicht, ihre Indexierungs-Präferenzen in einem besonderen Formular anzulegen, das bei der Einrichtung des blog-services auszufüllen ist und den Webmaster anweist, wie er seine eigene robots.txt-Datei konfigurieren soll (http://buzz.blogger.com/2012/03/customize-your-search-preferences.html).

[11] Diese Empfehlung ist besonders relevant in dynamischen Umgebungen oder auf komplexen Webseiten, wo die robots.txt-Lösung nicht ausreichend mit der Größe der Webseite skalieren könnte. Ein Beispiel der Nutzung der robots.txt-Kommandos, um einer Suchmaschine das Verfallsdatum einer Seite zu signalisieren, ist verfügbar unter http://googleblog.blogspot.fr/2007/07/robots-exclusion-protocol-now-with-even.html. In gleicher Weise signalisiert die sitemap.xml-Datei, wie oft sich eine

- Besondere Sorgfalt sollte beim Schreiben der robots.txt-Datei im Bezug auf die lexikalische und semantische Korrektheit der Anweisungen wie auch ihrer inhärenten logischen Konsistenz gewidmet werden (um gegensätzliche und/oder überlappende Anweisungen zu vermeiden). Es sollte betont werden, dass ein Crawler in Ermangelung *spezifischer Ausschluss-Anweisungen* in der robots.txt-Datei annehmen wird, dass der Administrator die Indexierung der Website oder die Indexierung bestimmter Unterverzeichnisse gestattet (d. h. ein Crawler wird annehmen, dass der Inhalt der Website für Suchmaschinen verfügbar gemacht werden soll).

- Es sollte beachtet werden, dass das robots.txt-Protokoll nicht für die Regelung des Zugriffs auf besonders „riskante" Inhalte wie Verkehrsdaten elektronischer Kommunikationsdienste, Inhalte von SMS-Nachrichten, Speicher von Anrufbeantwortern, Aufenthaltsdaten, Finanzdaten etc. geeignet, noch dass es zur Verhinderung des Zugangs zu spezifischen administrativen Bereichen einer Website gedacht ist. Das robots.txt-Protokoll ist kein Ersatz für Verschlüsselung oder Zugriffskontrollmechanismen.

- Wenn ein Webmaster zu signalisieren beabsichtigt, dass bestimmte Seiten und/oder Dateien nicht von Suchmaschinen indexiert werden sollen, sollte besondere Sorgfalt auf die Auswahl der URLs verwendet werden. Tatsächlich könnte, da die robots.txt-Datei öffentlich sichtbar ist, das Vertrauen auf „selbsterklärende" URLs letztlich die Verfügbarkeit der betreffenden Inhalte erhöhen und damit die Vorteile des Protokolls zunichtemachen. Der Inhalt der robots.txt-Datei ist für Hacker besonders wertvoll, wie auch für jede andere Instanz, die versucht, personenbezogene Daten zu verbreiten oder zu beschaffen.

5. Empfehlungen für Suchmaschinen

Als eine ihrer Kernaktivitäten arbeiten Anbieter von Suchmaschinen überwiegend als Informationsvermittler/Intermediäre[12]. Allerdings gibt es auch bestimmte Arten der Verarbeitung, für die sie als eigenständige verantwortliche Stellen agieren.

Insbesondere führen einige Suchmaschinen viele verschiedene Aktivitäten durch, die von der Indexierung von Webseiten bis zur zeitweisen Speicherung des diesbezüglichen Inhalts reichen, um Nutzern das Auffinden der Informationen in Fällen zu ermöglichen, in denen ein Server und/oder Link abgeschaltet/nicht verfügbar ist. Dieses „Caching" stellt eine Wiederveröffentlichung dar, für die der Anbieter der Suchmaschine als verantwortliche Stelle betrachtet wird[13].

Dementsprechend werden die folgenden Empfehlungen für Anbieter von Suchmaschinen unterschieden im Hinblick auf die unterschiedlichen Rollen, die sie spielen.

Webseite verändern kann, und die Priorität, die ein Webmaster einer URL beimisst, was der Suchmaschine erlaubt, die angemessene Auffrischungsgeschwindigkeit zu wählen. Vgl. auch http://lists.w3.org/Archives/Public/public-privacy/2012OctDec/0224.html

[12] Vgl. die Stellungnahme 1/2008 zu Datenschutzfragen im Zusammenhang mit Suchmaschinen (WP 148) der Artikel-29-Datenschutzgruppe der Europäischen Datenschutzbeauftragten (http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2008/wp148_de.pdf). Man beachte, dass diese Angelegenheit gegenwärtig vor dem Europäischen Gerichtshof verhandelt wird.

[13] Wie in der Stellungnahme zu Suchmaschinen von der Artikel-29-Datenschutzgruppe der europäischen Datenschutzbeauftragten (WP 148) betont wird, ist „... jegliche Zwischenspeicherung von auf indexierten Webseiten enthaltenen, personenbezogenen Daten über diesen aus Gründen der technischen Verfügbarkeit notwendigen Zeitraum hinaus (...) als eine unabhängige Neuveröffentlichung anzusehen. Nach Auffassung der [Artikel-29-] Arbeitsgruppe liegt die Verantwortung für die Einhaltung der Datenschutzgesetze hier beim Anbieter derartiger Caching-Funktionalitäten in seiner Rolle als Verantwortlicher für die Verarbeitung der personenbezogenen Daten, die in den zwischengespeicherten Veröffentlichungen enthalten sind."

Bloße Indexierung

- Suchmaschinen sollten die von den Websites in Bezug auf die Inhalte, die sie enthalten ausgedrückten Präferenzen immer respektieren, sei es durch die robots.txt-Datei oder durch andere „noindex“-Markierungsmechanismen, einschließlich Anweisungen zu Verfallsdaten. Solche Indexierungs-Präferenzen können vor der ersten Durchsuchung der Website ausgedrückt werden, oder nachdem sie schon durchsucht worden ist. Im letzteren Fall sollten Aktualisierungen der von einer Suchmaschine durchgeführten Indexierung so schnell wie möglich durchgeführt werden.

- Suchmaschinen sollten die Effizienz ihrer Kommunikationskanäle mit Webmastern erweitern, um schnell über jegliche Veränderung der Indexierungs-Präferenzen in Kenntnis gesetzt zu werden, die von Webmastern durch die geeigneten Anweisungen des robots.txt-Protokolls ausgedrückt werden, oder von jeder Veränderung von Objekten innerhalb einer Website. Die Aktualisierungs/Berichtigungs-Prozeduren sollten so datenschutzfreundlich wie möglich sein – insbesondere sollten keine zusätzlichen personenbezogenen Daten von Nutzern verlangt werden, die verlangen, dass bestimmte personenbezogene Daten aktualisiert/berichtigt werden.

- Suchmaschinen sollten ihre Crawling-Häufigkeit den Suchpräferenzen der Webmaster anpassen. Sie sollten auch jegliche Anforderungen von Webmastern zur Re-Indexierung ihrer Webseiten oder von Teilen davon infolge der Löschung oder Berichtigung von personenbezogenen Daten unverzüglich ausführen.

- Da es bisher keine konsistente Interpretation der in einer robots.txt-Datei oder anderen Signalisierungsmechanismen für Indexierungspräferenzen (z. B. Metatags, sitemap.xml.-Dateien) enthaltenen Anweisungen durch Suchmaschinen gibt, ist schwer vorherzusagen, welchen Einfluss solche Mechanismen auf die Indexierung einer Website durch die verschiedenen Crawler haben wird. Es ist wünschenswert, dass sich Suchmaschinen in dieser Hinsicht auf einen „modus operandi“ einigen. Die für die einzelnen Befehle anwendbaren Mechanismen sollten in klarer Weise auf einer Seite beschrieben werden, auf die von Nutzern leicht zugegriffen werden kann (z. B. von den Hauptseiten des Suchmaschinenportals).

- Suchmaschinen sollten in einem größeren Maße in die Unterstützung von Website-Administratoren eingebunden sein, indem sie Anleitungen und/oder Werkzeuge für die automatisierte Analyse von Indexierungs-Präferenzen zur Verfügung stellen. Dies wird Administratoren ermöglichen, zu überprüfen, welche Effekte die von Ihnen gegebenen Befehle in Bezug auf die Indexierung haben werden.

- Suchmaschinen sollten die Terminierung und Kriterien des „crawling“, das sie auf einer bestimmten Website durchführen, klarer spezifizieren, so dass Administratoren und Nutzer in vernünftiger Weise abschätzen können, wie lange eine bestimmte Information als Suchergebnis verfügbar bleibt.

Zeitweise Speicherung von durchsuchten Informationen

- Suchmaschinen sollten spezifische Crawler implementieren, wenn sie beabsichtigen, Daten nach verschiedenen Kategorien und für verschiedene Zwecke (z. B. generelle Indexierung, Nachrichten, Bilder, etc.) zu gruppieren, um Administratoren von Webseiten zu ermöglichen, den Kontext, in dem Informationen veröffentlicht werden, besser zu kontrollieren.
- Bei der Indexierung einer Website sollten Suchmaschinen komplexere und granularere Instruktionen für ihre Crawler zulassen, wie beispielsweise die folgenden:

  - Die Erlaubnis zur Indexierung von Informationen für bestimmte Zwecke (z. B. Allzweck-Suchmaschinen vs. Nachrichten-Suchmaschinen, etc.)[14];
  - die Erlaubnis, Informationen zeitweise für bestimmte Zwecke zu speichern, einschließlich diesbezüglicher Zeitbegrenzungen (z. B. caching, snippets);
  - die Erlaubnis, Informationen für bestimmte Zwecke an Dritte weiterzugeben;
  - die Erlaubnis, die abgefragten Informationen für bestimmte Anwendungsfälle[15] basierend auf dem Vorkommen von Eigenschaften, wie geografischer Lage oder IP-Adressräumen zu verarbeiten.

- Wo die Durchsuchung eine zeitweisen Speicherung von Inhalten einer Website für andere Zwecke zur Folge hat, als es Nutzern zu ermöglichen, auf diese Inhalte im Falle zuzugreifen, dass der betreffende Server/das betreffende Netzwerk abgeschaltet/nicht verfügbar ist, sollten Suchmaschinen die Administratoren von Websites mit eindeutigen, spezifischen Informationen über den Zeitablauf versorgen und über technische Mechanismen, die für diese Speicherung gelten.

- Suchmaschinen sollten aufgrund spezifischer Anforderungen von Webmastern durch deren Such-Präferenzen jegliche Cache-Kopie der von Webseiten abgerufenen Daten unverzüglich löschen, und sollten von der weiteren Verarbeitung dieser Daten absehen, um das Risiko der Verbreitung der Daten und deren übermäßiger Exponierung zu begrenzen.

6. Ein abschließender Vorbehalt

In diesem Papier hat die Arbeitsgruppe Werkzeuge für die Kontrolle der Verfügbarkeit (personenbezogener) Daten im Web untersucht, die heute für Nutzer, Webmaster und Suchmaschinen verfügbar sind, zumeist gegründet auf die Begrenzung der Verfügbarkeit von Inhalten auf einer Website entweder durch Anwendung von (automatisierten) Löschmechanismen[16] oder durch die Implementierung von Protokollen zur Signalisierung von Suchpräferenzen. Es sollte daran erinnert werden, dass letztere immer noch auf einfachen ein/aus- (binären) Regeln für Crawler beruhen, die vor über 15 Jahren entworfen wurden. Im Gegenzug sind Suchmaschinen über die Jahre immer komplexer geworden und der ehe simple Inklusions-/Exklusionsmechanismus, der dem betreffenden Protokoll zugrunde liegt, ist nicht länger vollständig fähig, das fortwährend wachsenden Ausmaß der Gewinnung und Speicherung von Daten zu bewältigen. Es sollte z. B. herausgestellt werden, dass die Ver-

[14] Vgl. z. B. die von der italienischen Kartellbehörde verlauteten Feststellungen infolge einer Beschwerde der italienischen Vereinigung der Zeitungsverleger gegen Google. Danach verpflichtete sich Google öffentlich auf eine Reihe von Maßnahmen, um Verlage mit Werkzeugen auszustatten, die ihnen dabei helfen sollen, zwischen der Indexierung von Inhalten auf der allgemeinen Suchmaschine und der Indexierung auf der Nachrichten-Suchmaschine zu unterscheiden.

[15] Wegen der zunehmend komplexen Anwendungsfälle, die auf die von Suchmaschinen durchsuchten Informationen anwendbar sind, könnte es angemessen sein, das gegenwärtige Muster umzudrehen, nachdem Crawler eine Information lesen dürfen, wenn eine Anweisung formal inkorrekt ist oder von dem Crawler nicht interpretiert werden kann. Wenn es sich als unmöglich erweist, eine komplexe Menge von Anweisungen zu interpretieren, sollte dies automatisch als ein Verbot der Indexierung/Speicherung durch den Crawler interpretiert werden.

[16] Es ist hervorzuheben, dass aufgrund der öffentlichen Natur des Web andere Zugriffskontrollmechanismen wie die Authentifizierung von Nutzern und/oder Verschlüsselung von Daten implementiert werden sollten, wenn der Administrator einer Website Inhalte aus der „Öffentlichkeit“ entfernen möchte.

fügbarkeit von Daten (einschließlich Daten, die Nutzer über sich selbst preisgeben), in Kombination mit Gesichtserkennungstechniken und Aufenthaltsinformationen, letztendlich eher die Indexierung von Individuen als nur von Inhalten oder Informationen ermöglichen kann. Ein vordringliches Augenmerk auf diese Aspekte ist deswegen notwendig.

Ein anderer, zukünftiger technologischer Durchbruch für den besseren Schutz personenbezogener Daten im Web könnte die Entwicklung des "policy-aware, semantic Web" sein, in dem Daten untrennbar mit Attributen (z. B. einer "Bedeutung") und Zugriffsregeln verknüpft werden könnten. Dies würde auf der einen Seite die Schaffung von neuen Beziehungen zwischen Daten ermöglichen und das Konzept einer vernetzten Welt erweitern, und auf der anderen Seite effektivere Mechanismen zur Erkennung und Auffindung von Inhalten ermöglichen, und potenziell auch von Kopien von Informationen, die mit diesem Objekt in Beziehung stehen, gestützt auf den Abgleich von Attributen (anstatt auf einfache Textvergleiche, wie dies heute stattfindet). Dies macht es vorstellbar, Informationen von einer Vielzahl von Websites zu entfernen und Suchergebnisse von Websites zu entkoppeln, und damit jegliche unbeabsichtigte Verbreitung von Daten zu vermeiden[17].

Natürlich sollte die Benutzbarkeit des Web nicht unterminiert werden und es muss eine Balance zwischen Innovation und den Grundrechten des Individuums auf Datenschutz und Schutz der Privatsphäre gefunden werden. Die Möglichkeit der Einführung granularerer Mechanismen, die nicht auf der einfachen Exklusions-/Inklusionsregel basieren, sollte weiter bedacht werden, aber eher darauf gerichtet sein, Betroffene in die Lage zu versetzen, ihre eigenen Suchpräferenzen besser auszudrücken und die Information mit dem angemessenen Kontext zu verbinden (z. B., indem es Betroffenen ermöglicht wird, zu signalisieren, ob eine bestimmte Information noch aktuell oder relevant ist, oder das Vorkommen jeglicher Ereignisse, die Auswirkungen auf diese Informationen gehabt haben könnten). Dies würde den Betroffenen mehr Möglichkeiten eröffnen als die einfache Wahl zwischen pauschaler, überschießender Verfügbarkeit im Web oder einem kompletten Verzicht auf neue Technologien.

Es gibt bedeutende und wachsende ökonomische Interessen sowohl bei Suchmaschinen als auch bei Administratoren von Websites, die auf die größtmögliche Verfügbarkeit von Daten durch die Implementierung der Indexierung von Daten und Informationen dringen. Diese Indexierung von Websites dient den ökonomischen Interessen bestimmter Marktteilnehmer und die Entfernung öffentlich zugänglicher Webinhalte oder die Signalisierung, dass solche Inhalte nicht durch eine Suchmaschine indexiert und abgerufen werden sollten, wird zwangsläufig Auswirkungen auf Geschäftsmodelle und die Marktdynamik haben. Eine Zusammenarbeit der verschiedenen Interessenvertreter ist notwendig, um die diesbezüglichen Interessen mit der Notwendigkeit des Schutzes der Privatsphäre angemessen in Einklang zu bringen.

[17] Google hat kürzlich eine Aktualisierung seines Suchalgorithmus angekündigt, die die Platzierung von Sites mit einer großen Anzahl von Mitteilungen über Löschungen herunterstufen wird und die nur in Fällen der Verletzung von Urheberrechten angewandt werden soll (http://insidesearch.blogspot.fr/2012/08/an-update-to-our-search-algorithms.html oder http://www.google.com/insidesearch/howsearchworks/).