Datenschutz im Web 2.0

Tobias Lange

http://www.einbecker.net/datenschutz_im_web_2.0/

3. Juni 2007

Inhaltsverzeichnis

0 Zusammenfassung

Das *Web 2.0* als *soziale*, weil nutzerorientierte Form des Internets bietet neue, offen zugängliche Kommunikationsformen, die schwere Folgen für den Datenschutz haben. Die Nutzer sind bereit, weite Teile ihres *Onlinelebens* öffentlich zu machen. Dieses Papier stellt die veränderten Bedingungen des Datenschutzes in dieser neuen

Welt dar.

0.1 Lizenz

1 Einleitung

1.1 Übersicht

1,1 Milliarden Menschen[1] benutzen das Internet, Tendenz steigend. In zunehmenden Maße werden dort soziale Beziehungen abgebildet – oder sogar dorthin verlagert. Die Nutzer geben freiwillig viele persönliche Daten an Unternehmen, die diese mit Daten anderer Dienste vernetzen und weiterverarbeiten. Diese parallele Entwicklung von technischen Möglichkeiten und sozialer Vernetzung wird unter dem Schlagwort *Web 2.0* zusammengefasst, welches die revolutionäre Bedeutung ausdrücken soll, den dieser Wandel für die Nutzer ausmacht. Durch die zunehmende Veröffentlichung dieser Informationen entstehen allerdings Profile, die weiten Einblick in die Persönlichkeiten der Nutzer zulassen.

1.2 Der Begriff Web 2.0

Web 2.0 ist ein Schlagwort, das nicht genau definiert werden kann. Es bezeichnet ein Konzept[2], das zum einen eine offene, standardisierte technische Basis, die die

[1] Stand: März 2007, http://www.internetworldstats.com/stats.htm

[2] siehe insbesondere Tim O'Reilly: "What Is Web 2.0?" (http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html)

gewünschte Informationsvernetzung ermöglicht, und zum anderen die Umsetzung von vernetzten Diensten auf dieser Basis umfasst.

1.2.1 Die technische Basis

Das *Web 2.0* basiert auf einfachen Technologien, die zum Teil schon Jahre vor der Einführung des Begriffs 2004 entwickelt und auch eingesetzt wurden, jedoch erst in der Kombination ihrer Möglichkeiten den Kern des *Web 2.0* bilden. Einen wichtigen Teil bilden für jedermann offene APIs auf XML-Basis, die es Anwendungswendungsentwicklern erlaubt, auf die Daten anderer Anwendungen zuzugreifen. Dies ermöglicht so genannte Mashups, das heißt die Kombination verschiedener Datenbestände zu einer neuen Anwendung. Ein gutes Beispiel hierfür bietet QYPE[3], welches Kartendaten von Google Maps[4] mit benutzergenerierten Kritiken zu Dienstleistern wie zum Beispiel Restaurants kombiniert.

Während kompliziertere Datenstrukturen eigene XML/SOAP[5]-basierte APIs anbieten, ist auf einfachen Seiten wie Weblogs, Nachrichtenseiten, Fotodiensten etc. das Anbieten der Daten in Syndikationsformaten (so genannten *Feeds*) wie Atom[6] oder RSS[7] verbreitet. Diese Formate sowie Zusätze wie hCard[8] ermöglichen einfache Semantiken, die maschinell leicht erfassbar sind.

Eine weitere Verwendung von XML auf vielen *Web 2.0*-Seiten ist AJAX[9], welches das dynamische, asynchrone Nachladen von Inhalten mittels Javascript und in XML formulierten Requestobjekten erlaubt. Dadurch wird das statische, synchrone Modell der Client-Server-Kommunikation durchbrochen und es entstehen dynamische, schnell antwortende Anwendungen innerhalb des Browsers.

[3] http://www.qype.com/
[4] http://maps.google.com/
[5] Simple Object Access Protocol, ein Protokoll, mit dem Daten zwischen Systemen ausgetauscht und Remote Procedure Calls aufgerufen werden können.
[6] definiert in RFC 4287: http://atompub.org/rfc4287.html
[7] Really Simple Syndication, mehrere ähnliche Feedformate in unterschiedlichen Versionen. Siehe Wikipedia: RSS (file format): http://en.wikipedia.org/wiki/RSS_(file_format)
[8] hCard ist ein sogenanntes Microformat (http://microformats.org/), bei denen semantische Zusätze in die Syntax des zugrundeliegenden Formats (HTML, RSS, etc.) eingebunden werden. hCard ist ein Visitenkartendaten-Microformat.
[9] Asynchronous Javascript and XML, Begriffseinführung durch Adaptive Path, http://www.adaptivepath.com/publications/essays/archives/000385.php

Weiterhin beinhaltet der Begriff *Web 2.0* Software, die die einfache Erstellung und Bereitstellung von Inhalten unterstützt. Klassische Beispiele hierfür sind Weblogsysteme und Wikis, allerdings auch Programme zum Bereitstellen von audiovisuellen Medien wie Video- und Audiopodcasts[10] oder aber Fotos. Normalerweise unterstützen diese Systeme Protokolle, die eine einfache Vernetzung der Inhaltsseiten mit anderen Seiten, auf die sie sich beziehen. Beispiele hierfür sind Benachrichtigungssysteme auf Weblogs (*Track*-[11] oder*Pingbacks*[12]) und die einfache Verlinkung innerhalb eines Wikis.

1.2.2 Die soziale Komponente

Auf dieser technischen Grundlage lassen sich also Anwendungen entwickeln, die leicht zu bedienen und stark vernetzt sind. Verbunden mit der Steigerung der Internetnutzung auf inzwischen 72 Minuten pro Tag[13] bedeutet dies ein Bedürfnis der Nutzer, ihren Lebensstil auch online zu verfolgen: Sie möchten ihre sozialen Beziehungen auch auf das Internet ausweiten, sie werden dort gepflegt oder auch erst aufgebaut.

Die vernetzte Struktur in Verbindung mit sozialen Aspekten, die in den Anwendungen vorhanden sind, begünstigt diese Entwicklung; Querverweise sind nicht nur Hyperlinks auf gewisse Dokumente, sondern auch Verknüpfungen zwischen den Personen hinter den Dokumenten. Dadurch übertragen sich soziologische Begriffe wie *Freundschaft* auf die virtuelle Welt, bei vielen Anwendungen werden Freundeslisten geführt und die Freundesanzahl angezeigt. Auch der Beziehungsstatus sowie mögliche partnerschaftliche Auswahlkriterien wie Größe, Gewicht, Haarfarbe oder aber auch das persönliche Einkommen können und werden von den Nutzern eingestellt.

Zusätzlich werden auch die Dokumente der Benutzer, wie beispielsweise Fotos, Videos, Weblog- oder Foreneinträge verknüpft, mit Schlagworten[14] versehen und

[10] Podcasts sind Feedformate, die Dateien verbreiten. Das wichtigste Beispiel hierfür sind Audiodateien für portable Medienspieler wie z. B. Apples iPod – daher der Name.
[11] Spezifiziert von Movable Type: http://www.movabletype.org/docs/mttrackback.html
[12] Spezifiziert von Ian Hickson: http://www.hixie.ch/specs/pingback/pingback
[13] TimeBudget 14, http://www.sevenonemedia.de/unternehmen/presse/pm/index.php?pnr=23231
[14] *tag* (Englisch): Schlagwort, das mit einer Information verknüpft wird. Siehe: http://en.wikipedia.org/wiki/Tags

kommentiert, so dass die Inhalte und die Nutzer immer stärker miteinander vernetzt sind.

1.2.3 Das Geschäftsmodell

Das *Web 2.0* wird in zunehmenden Maße als lukratives Geschäftsumfeld gesehen. Durch die Möglichkeiten, die es den Nutzern bietet, sind diese bereit, auf den Plattformen ihre eigenen Inhalte einzustellen. Als Anbieter muss man also nur noch eine Infrastruktur bereitstellen, die Anwender nutzen diese (so sie deren Bedürfnisse befriedigt), um die Inhalte selbst einzupflegen. Daher lassen sich mit relativ kleinen Unternehmen Dienste mit hohen Nutzerzahlen entwickeln, die zum Beispiel für Werbepartner interessant sind. So zahlte Google an MySpace ca. 900 Millionen Dollar[15], um für ca. 2,5 Jahre exklusiver Werbe- und Suchanbieter der Seite zu sein. Zusätzlich sinkt mit dem Preis für Internet-Datenverkehr[16] eine der Hauptkosten dieser Unternehmen kontinuierlich.

Unternehmen aus diesem Umfeld wurden zu durchaus hohen Preisen aufgekauft, wie folgende Beispiele zeigen: MySpace wurde für 580 Millionen US-$ von News Corp. akquiriert (ca. 10$/Mitglied), Google kaufte YouTube für 1,65 Milliarden Dollar (256 $/Mitglied), und Holzbrinck bezahlte für StudiVZ ca. 100 Millionen Euro (100 Euro/Mitglied).

Das Geschäftsmodell dieser Firmen basiert bis auf wenige Ausnahmen[17] alleine auf der Durchführung von Marketingmaßnahmen, sei es klassische Werbeanzeigen auf den Seiten des Dienstes oder aber die Weitergabe der Daten an andere Dienste – die Daten, die die Nutzer bereitstellen, sind jedenfalls Kernpunkt jeder Strategie. Dass es hierbei zu einem Konflikt zwischen den Interessen der Dienste und dem Datenschutz kommt, ist offensichtlich.

[15] http://www.stern.de/wirtschaft/unternehmen/unternehmen/:Online-Markt-MySpace-Google/567229.html

[16] In Europa von 2005 auf 2006: -22%, http://www.dri.co.jp/auto/report/telegeo/telegeogig07.htm

[17] Ein Beispiel hierfür ist Flickr (http://flickr.com/), das keinerlei Werbung schaltet und sich alleine durch Pro-Accounts (mit besserer Funktionalität) für 25$ pro Jahr finanziert.

1.3 Beispiele

1.3.1 Flickr

Flickr ist eine 2004 gestartete Fotocommunity, die sich durch Abonnements von im Leistungsumfang gesteigerten Accounts finanziert. Der Basisaccount, der (im Umfang begrenzt) alle Funktionalitäten von Flickr ermöglicht, ist kostenlos. Flickr bietet auf seinen Seiten Zugriff auf über zwei Milliarden Fotos.[18]

1.3.2 MySpace

Myspace ist ein Dienst, der sich an ein junges Publikum richtet. Nach Angaben von MySpace haben 174 Millionen Menschen ein Profil auf der Seite, können sich vernetzen und Nachrichten austauschen. MySpace bietet Musikbands die Möglichkeit, auf deren Seite Stücke zu Promotionszwecken hochzuladen und einen direkten Kontakt mit Ihrer Fanbasis zu pflegen, welche diese Musikstücke dann in ihren Profilseiten einbinden können.

1.3.3 YouTube

YouTube ist ein Videoportal, dass die Nutzer dazu animieren soll, kurze Filme (unter 10 Minuten) auf der Seite hochzuladen.

[18] http://www.techcrunch.com/2007/11/13/2-billion-photos-on-flickr/

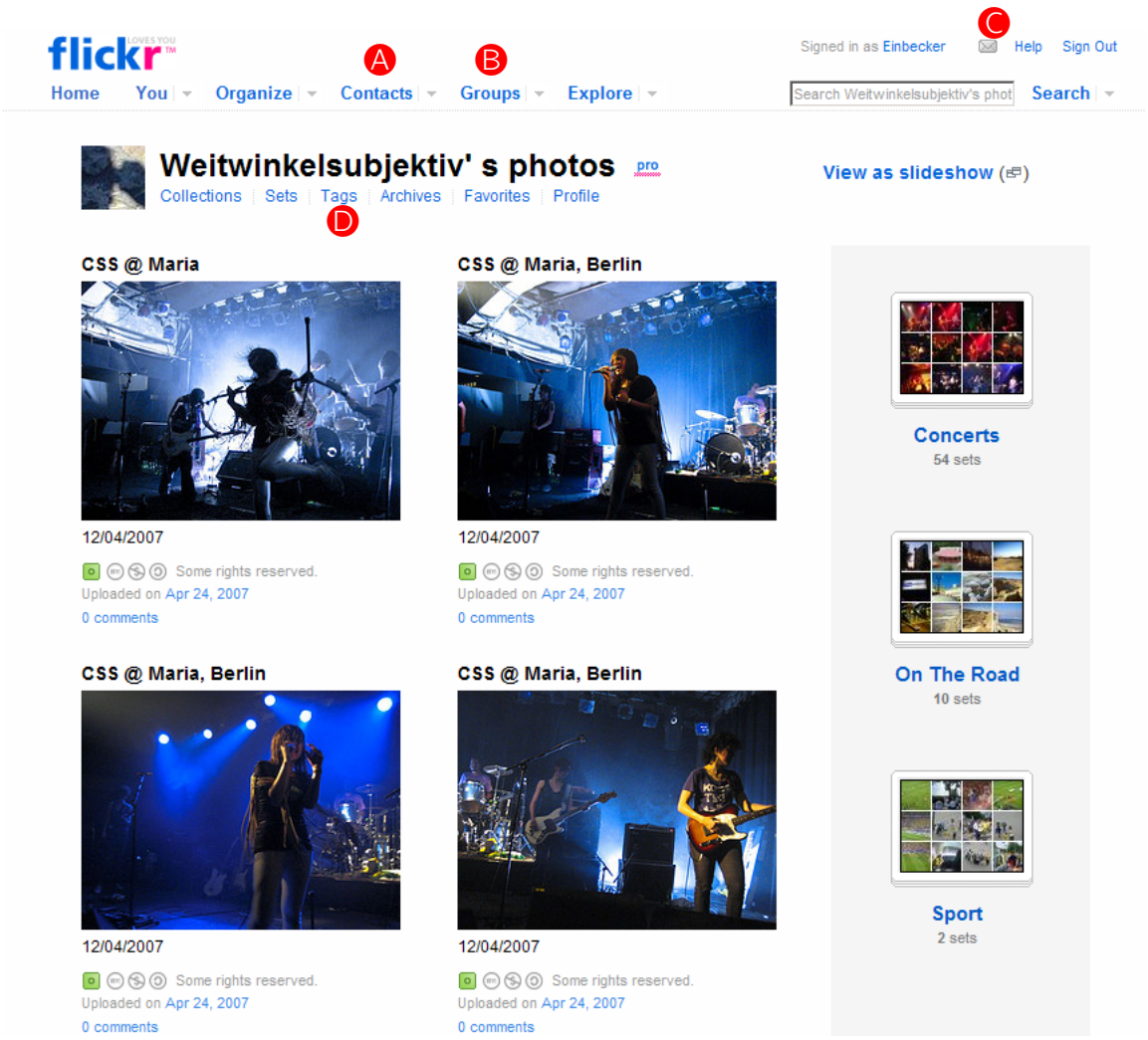

Abbildung 1: *Flickr.com* mit den üblichen Funktionen: (A) Kontaktverwaltung (sonst häufig Freundesliste genannt), (B) Gruppen, (C) private Nachrichten. Im aufgeräumten, von Weißraum dominierten Layout nimmt der Inhaltsbereich der Fotos den Hauptteil ein, am rechten Rand finden sich Ansammlungen von bestimmten Fotos zu einem gewissen Thema. Flickr macht extensiven Gebrauch von Tags (D), um Bilder zu katalogisieren und auffindbar zu machen. Auffällig ist das völlige Fehlen von Werbung, da das Geschäftsmodell anders als bei den meisten Anbietern auf Abonements setzt.

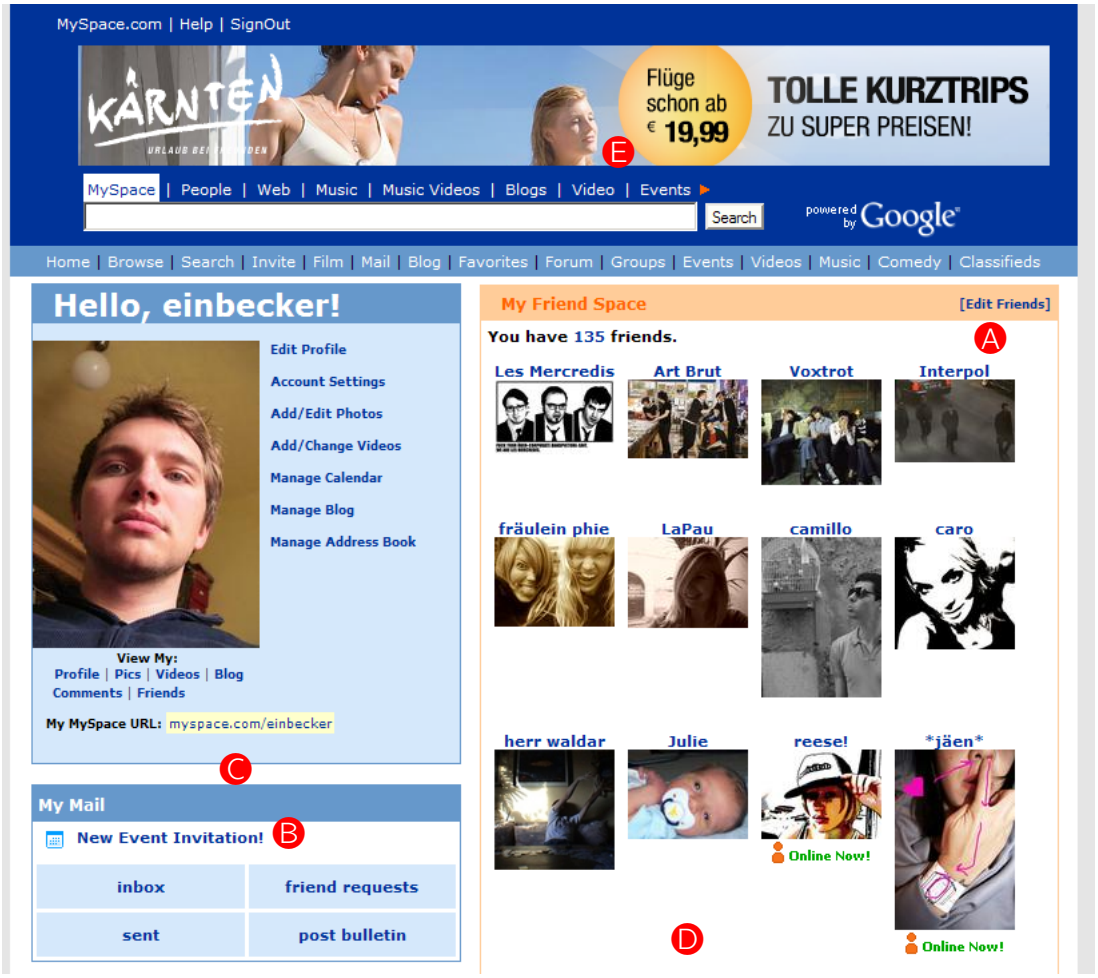

Abbildung 2: *MySpace.com* mit den Funktionen: (A) Freundesliste, (B) private Nachrichten, Geburtstage und Kalender. Das Layout wird hierbei von den unterschiedlichen Kontaktmöglichkeiten dominiert – der Austausch von sozialen Kontakten ist klares Hauptaugenmerk des Dienstes. An den Stellen (C) und (D) tauchen in normalen Layouts die eingebundene Musik sowie ein Kommentarbereich, eine öffentliche Pinnwand, auf. Bemerkenswert ist auch die große Werbung (E), die auf manchen Seiten auch noch größer sein kann, sowie die Nichtunterscheidung zwischen Musikbands und normalen Freunden in der Freundesliste.

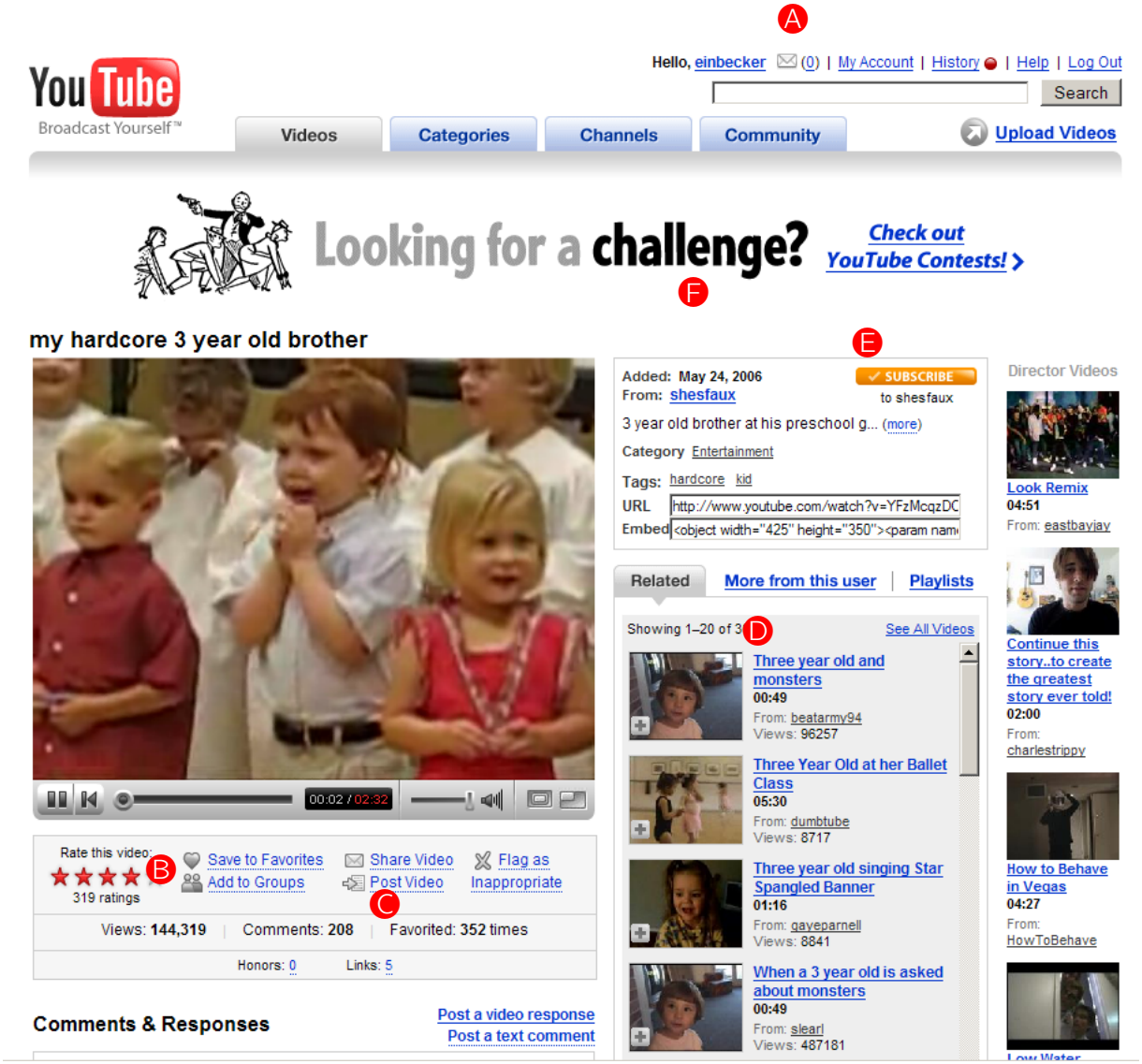

Abbildung 3: *YouTube.com* mit Nachrichtenverwaltung (A), Tagging/Bewertung (B) sowie Einbinde- und Empfehlungsmöglichkeit (C). Man beachte auch die verwandten Videos (D), die die Verweildauer auf der Seite erhöhen sollen, sowie die Werbung (F) – das Geschäftsmodell von YouTube. Es besteht eine Abonnementfunktion für Videos eines Urhebers (E), wobei diese nicht über RSS realisiert ist.

2 Datenschutz

2.1 Definition des Datenschutzbegriffs im Kontext des Web 2.0

Druch die ungenaue Definition des Begriffs *Web 2.0* ist die Eingrenzung der möglichen Anwendungen und damit auch der impliziten Bedingungen für den Datenschutz erschwert. Es fällt auf, dass die einzelnen Anwendungen jeweils nur ein geringes Datenschutzrisiko darstellen, die Vernetzung dieser jedoch zu weitaus größeren Risiken führt.

Zusätzlich erfordern die Anwendungen die teilweise Aufhebung des Schutzes personenbezogener Daten, da sie explizit dazu dienen, diese Daten an andere Personen weiterzugeben.

Aufgrund dieser Gegebenheiten lassen sich folgende Bedingungen definieren, nach denen sich die Anwendungen verhalten sollen:

- ▷ Die Anwendungen dürfen nur die für sie nötigen Daten erfassen.
- ▷ Die Weitergabe der Daten an andere Nutzer und Dienste muss dem Nutzer deutlich kenntlich gemacht werden. Idealerweise sollte die Weitergabe vom Nutzer selbst einstellbar sein.
- ▷ Pseudonymität gegenüber anderen Nutzern sollte gegeben sein. Gegebenenfalls sollten eindeutige Bedingungen definiert werden, die es dem Dienstanbieter erlauben, die Pseudonymität aufzuheben.
- ▷ Die gesamte Kommunikation, geringstenfalls jedoch der Loginvorgang und die Übertragung personenbezogener Daten, muss über sichere Protokolle erfolgen, die nicht durch Netzwerkanbieter oder andere Unbeteiligte abgehört werden können.
- ▷ Der Nutzer muss die volle Kontrolle über seine Daten haben: Sämtliche von ihm eingestellten Daten sollten auch aus dem System exportier- und löschbar sein. Einzig aus rechtlichen Gründen kann diese Bedingung, so fern nötig, eingeschränkt werden, so dass diese Daten zum Beispiel auf Anordnung eines Gerichtes, wiederhergestellt werden können.

- ▷ Der Nutzer hat die volle Kontrolle über die Vernetzung seiner Daten: Verknüpfungen innerhalb des Dienstes zu seinen Daten sollten in seiner Kontrolle liegen.

2.2 Angreifermodell und -möglichkeiten

Das Angreifermodell ist ähnlich ungenau wie die Definitionen des Begriffs des *Web 2.0* sowie des Datenschutzes hierin. Grundsätzlich sind zwei Szenarien unterscheidbar:

Zum einen die wahllose Sammlung von allen Daten zu allen Personen, um so Informationen über eine gewisse (große) Gruppe zu sammeln, so zum Beispiel die Gesamtheit der Nutzer einer Plattform oder alle Blogger in einem bestimmten Land. Bei dieser Art des Angriffs erfolgen die Zugriffe automatisch von einem oder mehreren Rechnern, der Angreifer selbst ist dabei als passiv zu klassifizieren, auch wenn die genutzten Programme durchaus auch aktiven Charakter haben können, beispielsweise zur Überwindung von Challenge-Response-Aufgaben wie Captchas[19]. Die Informationen selbst werden gelesen und gespeichert. Hierbei ist wichtig, dass eine spätere Filterung der riesigen gewonnenen Datenmengen durch die oben besprochene durch den Nutzer erzeugte Semantik stark vereinfacht ist.

Die andere Möglichkeit ist die gezielte Sammlung von Informationen zu einer Person. Hierbei kann der Angreifer aktiv eingreifen und den Angriff sogar komplett manuell durchführen, da er genau weiß, nach welchen Informationen er zu suchen hat.

Beiden Modellen gemein ist, dass der eigentliche Angriff nicht illegal oder unter Umgehung von Sicherheitsmaßnahmen erfolgt: Es handelt sich um öffentliche oder semiöffentliche[20] Daten, die vom Angreifer ausgelesen werden. Die Nutzer sind sich jedoch nicht bewusst, dass ihre Daten für jeden verfügbar sind oder halten einen solchen Angriff für sehr unwahrscheinlich, weshalb sie ihn billigend in Kauf nehmen.

Nicht eingegangen werden soll hierbei auf die gut untersuchte Problematik der Klartextübertragung von Daten mittels HTTP: Es sollte klar sein, dass über eine

[19] In Bilder eingebette Zeichenfolgen, die vom Anwender eingegeben werden müssen, um zu beweisen, dass er menschlich ist. Siehe auch weiter unten.

[20] Nur den Nutzern eines Dienstes zugängliche

solche HTTP-Verbindung übertragene Daten zusätzlich zu den hier beschriebenen Methoden auch noch über die HTTP-Pakete gewonnen werden können.

Auch sollte in diesem Zusammenhang noch auf die zunehmende Nutzung von öffentlicher Infrastruktur wie offenem WLAN in Cafés durch die *digitale Bohème*[21] hingewiesen werden, was natürlich weitere Datenschutzprobleme verursacht.

Die Möglichkeiten der Angreifer sollten als potenziell sehr weitreichend angesehen werden: Nicht nur Werbeunternehmen (oft mit hohen Budgets in diesem Bereich) interessieren sich für die gewonnenen Daten, auch Regierungen von vielen Staaten nutzen diese.[22][23][24]

2.3 Pseudonymität

Durch den Wunsch der Nutzer, von anderen Nutzern wiederkannt zu werden, erschwert sich die Anonymität durch Pseudonyme. Prinzipiell könnten bei den meisten Anwendungen durchaus Transaktions-, jedoch mindestens dienstgebundene Pseudonyme verwendet werden. Da die Nutzer jedoch die Verkettbarkeit ihrer Person über die verschiedenen Dienste wünschen, bleiben nur noch Rollen- oder Personenpseudonyme übrig, die nur eine geringe Anonymität bieten.

Allerdings gibt es Beispiele von Weblogautoren, die trotz einer hohen Bekanntheit und einer weiten Verknüpfung ihres Profils auf vielen Seiten geschafft haben, für den Großteil der Nutzer anonym zu bleiben. Dies setzt aber einen sorgsamen Umgang mit den personenbezogenen Daten voraus – sind die Daten einmal in einem Dienst vorhanden, lassen sich diese von allen sehr schnell nachvollziehen.

---

[21] Sascha Lobo und Holm Friebe: Wir nennen es Arbeit, http://www.wirnennenesarbeit.de/. Die *digitale Bohème* ist eine Gruppe urbaner Nutzer des Web 2.0, deren Lebensmittelpunkt und auch Arbeitsleben sich auf diesen Bereich verschiebt

[22] Amerikanische Nachrichtendienste benutzen das *Web 2.0*, um Daten zu gewinnen: "NSA looking at social-networking spaces" (http://www.physorg.com/news69605300.html)

[23] Die Ägyptische Regierung nahm im Mai 2006 Aktivisten fest, unter denen sich auch Blogger befanden: "[...], the government has clamped down violently on protesters, arresting hundreds of activists including Alaa and a handful of other bloggers." (PBS Mediashift, http://www.pbs.org/mediashift/2006/05/digging_deeperblogs_wiki_googl.html)

[24] News.com: "Dictatorships catching up with Web 2.0" (http://news.com.com/2010-1028_3-6155582.html)

2.4 Vergleich mit dem alten Web

Im World Wide Web herkömmlicher Art bestanden Beziehungen nur zwischen Nutzern und Diensten und, getrennt hiervon, von Diensten mit anderen Diensten, wobei letztere Verbindungen keine Daten über die jeweiligen Nutzer enthielten. Die meisten Dienste, die dort zum Einsatz kamen (Beispielsweise E-Mail-Anbieter, Webshops, Foren etc.) haben in ihrem abgeschlossenen Raum die Daten der Nutzer benutzt, diese flossen jedoch nicht zu anderen Dienstanbietern und waren auch nicht öffentlich einsehbar, weder als HTML-Seite noch mittels Zugriff über eine API. Eine Verkettung, beispielsweise über die E-Mail-Adresse, war zwar möglich, aber nur, wenn die jeweiligen Dienstanbieter miteinander kooperierten. Dies war generell eher ungewöhnlich, da sich diese grundsätzlich als Konkurrenten betrachteten.

In der neuen Ausprägung des WWW trifft dies alles nun zu. Die meisten Dienstanbieter öffnen ihre Plattformen, lassen auch Nichtmitglieder auf Seiten zugreifen und bieten anderen Anbietern APIs an, damit diese ihre Datenbestände nutzen können. Dies ermuntert die Nutzer dazu, selbst Verknüpfungen herzustellen – die Einbindung von Wunschlisten und Videoinhalten sind hier nur zwei Beispiele. Dienstanbieter sowie externe Dritte sind also generell jederzeit in der Lage, Verknüpfungen zwischen den Pseudonymen auf unterschiedlichen Plattformen herzustellen.

2.5 Identitätsmanagement

Das Management der eigenen (Online-)Identität(en) wird zukünftig ein wichtiges Thema sein.[25] Allerdings ist das größte Problem, was die meisten Nutzer mit den *Web 2.0*-Diensten haben, nicht der Datenschutz, sondern die Notwendigkeit des mehrfachen Logins bei allen Diensten, die ein Nutzer benutzt. Die Lösung dieses Problems wird allgemein als Single Sign-On (SSO) bezeichnet.

Daraus folgt auch, dass der von vielen modernen Webdiensten favorisierte Identitätsmanagementmechanismus OpenID ist, welches ursprünglich nur zur Lösung

[25] Mario Sixtus: „Das nächste große Ding im Web trägt den Namen Identität. Und darüber wird noch sehr, sehr viel diskutiert und gestritten werden." (http://www.sixtus.net/entry/845_0_1_0_C/)

des SSO-Problems entworfen wurde. Viele andere Datenschutzprobleme werden von dieser Lösung gar nicht oder nur schlecht adressiert.[26]

So werden beispielsweise die Daten der benutzten Dienstanbieter (bzw. der Authentifizierung bei diesen) alle beim Identitätsanbieter gespeichert. Zusätzlich bietet OpenID standardmäßig keine Möglichkeit an, mehrere getrennte Identitäten zu benutzen.

Andere Anbieter, die versuchen, in diesem Bereich Lösungen anzubieten, sind SAML, Liberty, Microsoft Cardspace, PRIME und Credentica, wobei viele davon den von Datentschützern an diese gestellten Kriterien nicht erfüllen.[27]

Auch wenn Identitätsmangement ein Zukunftsthema ist und auch im *Web 2.0* durchaus auch von Interesse, wird sich im Rahmen dieses Papiers nicht weiter damit beschäftigt, denn das obige Kapitel über die Pseudonymität der Nutzung hat gezeigt, dass auch das beste System nicht helfen kann, wenn die Nutzer selbst die Verkettung über verschiedene Dienste wünschen.

2.6 Soziale Netzwerke

Unter sozialen Netzwerken versteht man ursprünglich die Verbindungsstruktur von Menschen mit anderen Menschen, so stellt ein Freundeskreis beispielsweise ein solches soziales Netzwerk dar. Im Sinne des *Web 2.0* versteht man darunter eine Anwendung, die diese Definition mit neuen Technologien auf das Internet abbildet. Durch das Hinzufügen von Personen zu den *Freunden* oder *Kontakten* werden diese zu für die jeweilige Person zu einem Freundesbaum (die verschiedenen Ebenen repräsentieren *Freundschaftsgrade*: Freunde, Freunde von Freunden etc.) und insgesamt zu einem sozialen Netzwerk aufgebaut.

[26] Siehe insbesondere Ralf Bendrath: "OpenID – next big thing with lots of problems." (http://bendrath.blogspot.com/2007/04/openid-next-big-thing-with-lots-of.html)

[27] Fred Carter des Office of the Information and Privacy Commissioner of Ontario bezüglich der Erfüllung der *seven laws of identity*, aufgestellt von Cardspace-Architekten Kim Cameron: "The list was short, just PRIME and Credentica – note the absence of CardSpace. So, I just had to ask: 'does this mean that you believe CardSpace does not obey the seven laws?'. His reply? 'Yes.' Chris Bunio, a Senior Architect for Microsoft, was present. He did not dispute the claim." (Ben Laurie, http://www.links.org/?p=227) Siehe auch: http://bendrath.blogspot.com/2007/05/cardspaces-privacy-problems-now.html

2.6.1 Aufbau

Der Aufbau von sozialen Netzwerken im Internet ist grundlegend immer identisch:

- ▷ Freundeslisten und die Möglichkeit, Personen als *Freunde* zu markieren
- ▷ Nachrichtendienst (Privat und auf einer *Pinnwand*)
- ▷ Gruppierungen von Nutzern.

Zusätzlich versuchen die Betreiber noch, Alleinstellungsmerkmale herauszuarbeiten, die dem Nutzer einen Zusatznutzen bieten, um diese an die Plattform zu binden. Viel wichtiger ist jedoch meistens die Aufmachung bzw. das Branding an die jeweilige Zielgruppe, also die Anpassung von Design, Sprache, Funktionalität und Bedienerführung an die jeweiligen Nutzer.

2.6.2 Zielgruppen

Soziale Netzwerke sprechen unterschiedlich große Personengruppen an; manche versuchen, einen möglichst großen Markt abzudecken, wogegen andere sich sehr auf Nischensegmente konzentrieren und so ihr Marketing besser spezialisieren können. Die Nutzerzahlen sind sehr hoch, so gibt es zur Zeit mindestens 41 Dienste mit mehr als einer Million Mitglieder[28]; der größte Anbieter MySpace hat nach eigenen Angaben mehr als 174 Millionen Mitglieder.

Folgende Gruppen lassen sich ausmachen (mit jeweils großen Netzwerken, die speziell auf diese Gruppierungen abzielen):

- ▷ Musikinteressierte (Last.fm sowie auch MySpace)
- ▷ Studenten (Facebook, StudiVZ)
- ▷ Angestellte und Selbstständige (XING)
- ▷ Foto- und Videointeressierte (Flickr, YouTube)
- ▷ weitere spezielle Interessen- und Personengruppen.

[28] siehe Wikipedia: http://en.wikipedia.org/wiki/List_of_social_networking_websites

2.6.3 „Paralleluniversum“

Die Anbieter von sozialen Netzwerken haben eine parallele Welt zu den herkömmlichen etablierten Technologien im Internet geschaffen. E-Mails werden durch Nachrichten auf den jeweiligen Plattformen ersetzt, der Nutzverkehr läuft nicht mehr zwischen den einzelnen Nutzern beziehungsweise deren selbst ausgesuchten Dienstanbietern, sonder nur über die Infrastruktur der jeweiligen Plattform. Dadurch liegen dem Dienstanbieter alle Daten, die über diese Plattformen ausgetauscht werden, vor. Die Nutzer können meist gar keine Verschlüsselungs- und Anonymisierungstechnologien einsetzen, der Einsatz wird stark erschwert[29] , oder dies ist nicht mehr sinnvoll[30]. Die Daten liegen außerhalb des Einflussbereichs des Nutzers auf den Servern des Dienstanbieters.

Zusätzlich liegen die Daten der Nutzer im Gegensatz zu herkömmlichen HTML-Seiten in einer homogenen Datenstruktur vor, die teilweise auch schon semantisch aufbereitet ist. Dies ermöglicht es dem Anbieter und Dritten, diese Daten automatisch auszulesen.

Nur sehr selten erfüllen soziale Netzwerke auch nur einige der oben aufgeführten Datenschutzkriterien. Auch wenn durchaus kleine Fortschritte etwa im Bereich der sicheren Protokolle oder auch der Kenntlichmachung der Datenweitergabe bei einigen Diensten gemacht wurden, muss doch festgestellt werden, dass die Funktionsweise der Angebote neben den noch zu nennenden grundlegenden Probleme der Daten im *Web 2.0* eine weitere Gefahr für den Datenschutz darstellt.

Während bei anderen Diensten es dem Nutzer frei steht, für welchen er sich entscheidet, ist er bei der Wahl eines sozialen Netzes an diesen Dienst gebunden. Auch wenn es andere Dienste mit ähnlichen Funktionalitäten gibt, kann er nicht einfach wechseln, da seine Daten (Freundeslisten, Einträge, Nachrichten, etc.) fest mit dem gewählten Dienst zusammenhängen und man auch nur mit Nutzern des gleichen Netzwerkes kommunizieren kann.

[29] Beispielsweise das Verschlüsseln von Nachrichten durch Kopieren und Einfügen des zu verschlüsselnden Textes in PGP und darauffolgendes zurückkopieren.

[30] z.B. MIX-Einsatz bei gleichzeitigem Login auf einer Seite

2.6.4 Web 2.0-Dienste in Kombination als Ersatz für soziale Netzwerke

Es besteht die Möglichkeit, ein zentrales soziales Netz durch eine dezentrale Ansammlung von Applikationen des *Web 2.0* zu ersetzen. Die oben beschriebenen Funktionalitäten machen dies problemlos möglich: Die Benutzerseite wird durch ein Weblog ersetzt, das über eine Blogroll[31] verfügt, welches die Freundesliste ersetzt. Kommentare ermöglichen die öffentliche Kommunikation und ersetzen so die Pinnwandeinträge. E-Mails ersetzen private Nachrichten. Bilder, Audiodateien und Videos können von eigenen Seiten oder Dienstanbietern wie Flickr und YouTube eingebunden werden. Selbst Kalender lassen sich mit Hilfe von Diensten wie Upcoming.org, Last.fm oder auch dem Google Calendar veröffentlichen.

Das bedeutet, dass die zentralen sozialen Netzwerke nicht nötig sind, wodurch einige der obigen Datenschutzprobleme beseitigt werden können. Allerdings konzentrieren sich die Nutzer erfahrungsgemäß auf wenige Plattformen, so dass auch weiterhin durch die hohe Verknüpfungsdichte und semantische Aufbereitung ein automatisches Abgreifen der zur Verfügung gestellten Informationen möglich ist.

2.7 Verkettung über das eigene Blog

Wie schon erwähnt, wünschen sich die Nutzer eine Verknüpfung ihrer digitalen Identitäten. Dies geschieht sehr häufig über das eigene Blog: Der Anwender stellt auf seiner Seitenleiste Links zu bestimmten Anbietern ein, oder aber er nutzt Widgets[32] der Anbieter, um seine Präsenz auf der jeweiligen Plattform zu zeigen. Dies können zum Beispiel die letzten Fotos sein, die hochgeladen wurden, oder Musikstücke, die der Nutzer kürzlich gehört hat. Über diese Seite und die von den Browsern übertragenen Referer[33] lassen sich von allen Anbietern sämtliche Informationen automatisch gewinnen.

[31] eine Auflistung von verlinkten Weblogs

[32] Zusammensetzung aus Gadget (technisches Spielzeug) und Window: in Webseiten einbettbares Programm, dass gewisse Inhalte einer Plattform anzeigt

[33] verweisende Seite

2.8 Angriffsarten

Im folgenden möchte ich darlegen, welche Arten der Informationen sich durch *Web 2.0*-Dienste darstellen lassen, um festzustellen zu können, welche Informationen ein Angreifer gewinnen kann. Dabei ist, wie schon erwähnt, zu beachten, dass die hier vorgestellten Informationen nicht wirklich geschützt, sondern frei verfügbar sind, so dass im engeren Sinne nicht von einem *Angriff* gesprochen werden kann.

Grundsätzlich ist hierbei von einem passiven Angreifer auszugehen.[34]

2.8.1 Verkettungsmöglichkeiten bei einem Nutzer

Orte und Zeit Viele Dienstanbieter beziehen Orts- und Zeitinformationen in ihr Angebot mit ein. So bieten viele Fotocommunities die Möglichkeit an, Fotos zu *geotaggen*, also das Foto mit dem genauen Aufnahmeort zu verknüpfen. Auch manche Kamerahersteller benutzen schon GPS-Sensoren, um in die Fotos über EXIF[35] Ortsdaten vollautomatisch einzubetten. Ferner bieten Dienste wie Plazes[36] die Möglichkeit an, seine Loginpunkte in WLAN-Netze öffentlich anzuzeigen, um in Kontakt mit anderen Anwesenden zu treten. Zusätzlich fällt auch die Bewertung von Lokalitäten in diesen Bereich, wie sie Beispielsweise von QYPE[37] angeboten wird.

Die dabei bereitgestellten Informationen sind enorm: So werden komplette Fototouren durch Städte gemacht, bei denen alle paar Meter ein weiteres Foto mit Koordinaten geschossen wird. Durch zeitnahes Veröffentlichen (zum Beispiel mit Mobiltelefonen von der integrierten Kamera gemachten Aufnahmen) lässt sich hierbei jederzeit erfahren, wo sich der Nutzer gerade aufhält.

[34] Dies ist eigentlich nicht ganz korrekt: Der Angreifer erzeugt durchaus Nachrichten (in den Meisten fällen Requests an den jeweiligen Webserver), die ihn theoretisch identifizierbar machen. Einen solchen Angreifer aus den Logfiles heraus zu identifizieren ist allerdings so kompliziert, dass es praktisch unmöglich ist.

[35] Exchangeable Image File Format: In den Headerdaten von Bilddateien werden Metainformationen hinterlegt

[36] http://plazes.com/

[37] http://qype.com/

Beziehungen Die Verknüpfung mit anderen Personen entsteht klassischerweise über Freundeslisten, die das Friend-Of-A-Friend[38]-Prinzip umsetzen. Viele Dienste, insbesondere soziale Netzwerke, bieten die Eingabe eines Beziehungsstatuses an. Auch die Blogrolls sind eine Verkettung über Beziehungen zu anderen Menschen. Hier kommt zusätzlich oft XFN[39] zum Einsatz, welches eine semantische Verknüpfung zu den Personen (etwa: Freundschaft, Verwandtschaft, getroffen etc.) über HMTL-Attribute erlaubt.

Interessen Auch über die Interessen der Nutzer lassen sich Verkettungen bilden. Beispiele wären hier Wunschlisten beim Onlineversandhandel oder gehörte Musiktitel bei Musikseiten. Auch Produkte und deren Bewertungen fallen in diese Kategorie.

Reputation Reputation spielt eine große Rolle in der Definition der digitalen Identität. Klassische Reputationssysteme wie Bewertungsschemata für Nutzer spielen hier genauso eine Rolle wie Kommentare auf Weblogs oder Blogeinträge. Auch Verlinkungen und Freundeslisten sind Teil der Reputation.

Hier ist insbesondere zu beachten, dass aktive adaptive Angreifer noch einen weitaus größeren Schaden verursachen können: Durch fehlende Authentisierungsmöglichkeiten können diese die Reputation von anderen Personen übernehmen, um zum Beispiel einen Kommentar in deren Namen abzusetzen.

2.8.2 Mehrdimensionale Verkettung

Die vorgenannten Angriffsmöglichkeiten lassen sich auch problemlos in mehrere Dimensionen erweitern, so beispielsweise über Freundeslisten die Interessen des Freundeskreises oder auch deren Aufenthaltsorte lokalisieren. Auch wenn aus Datenschutzgründen dies sicherlich unerwünscht ist, ist es vom Nutzer genauso gewollt: Die Verkettung dieser Informationen sind für ihn der zentrale Nutzen, den diese Dienste bieten.

[38] FOAF
[39] XHTML Friends Network: http://gmpg.org/xfn/

2.8.3 Von Identitätsfetzen zu Profilen

Durch die vom Nutzer bereitgestellte Semantik lassen sich über diesen sehr leicht Informationen sammeln. Zusätzlich können Daten bei den beliebtesten Plattformen ohne größere Probleme ausgelesen werden, auch wenn keine formelle Semantik gegeben ist, da diese an der gleichen Position innerhalb eines HTML-Dokumentes eingebunden sind. So lassen sich aus den Summen dieser Informationsfetzen Profile erstellen, die nicht nur über einzelne Nutzer, sondern deren Freunde, Lokalitäten oder Interessen verkettet sind.

3 Schutzmöglichkeiten

3.1 technische Schutzmöglichkeiten

Der Schutz vor den beschriebenen Angriffen kann naturgemäß nur sehr beschränkt erfolgen, da die Daten generell öffentlich zugänglich sind. Abgesehen vom *verantwortungsvollen Nutzer*, der sich darüber bewusst ist, welche Informationen er über sich veröffentlicht, bestehen diese Mechanismen aus den folgenden:

Zugangskontrolle Nur bestimmte Nutzergruppen haben Zugriff auf die Daten. Dies können alle Nutzer des Dienstes, nur die Freunde, auch Freunde von Freunden (bis zu einem gewissen Grad) sowie jeweils neu kombinierte Gruppen sein. Hierzu muss sich der Nutzer authentisieren, was zum Beispiel über Login und Passwort erfolgen kann. Es ist zu beachten, dass dadurch die Anonymität des authentisierenden Nutzers aufgehoben ist. Auch kann schon ein einzelner Kooperierender aus der Gruppe die Vertraulichkeit der Daten aufheben.

Unterscheidung von Mensch und Maschine Um das automatische Auslesen durch Computerprogramme zu verhindern, wird versucht, gewünschte menschliche Nutzer von diesen zu unterscheiden. Hierzu gibt es verschiedene Ansätze:

- **Requestverhalten**: Die gesendeten Header an den Server werden untersucht, gegebenenfalls auch über einen gewissen Zeitraum. Dabei wird versucht, ty-

pische menschliche Verhaltensweisen von denen eines Programms zu unterscheiden.

- ▷ **Captchas**: Der Nutzer muss Daten aus einem Bild herauslesen. Hierbei werden Bilder genutzt, die für OCR nur schwer zu erkennen sind. Es können auch ähnliche Verfahren wie das Lösen einer Rechenaufgabe eingesetzt werden.
- ▷ **Fähigkeiten des Browsers**: Die Programme, mit den die Daten gesammelt werden, sind meist keine vollwertigen HTML-Browser und verstehen auch höchst selten JavaScript, so dass dadurch eingebundene Inhalte für das Programm nicht zugänglich sind.

Allerdings können alle diese Methoden auch den normalen Nutzer einschränken, zum Beispiel, weil er nur über eine geringe Sehfähigkeit verfügt oder sein Browser kein JavaScript unterstützt.

3.2 Gesellschaftliche Schutzmöglichkeiten

3.2.1 Gesetzgeber

Durch die Globalität des Internets ist es sehr schwierig, einheitliche gesetzliche Standards festzulegen, die den Schutz der Nutzerdaten regulieren. Selbst wenn eine Regulierung bestehen könnte, so hätte diese doch zwei Schwachpunkte:

- ▷ Sind die Daten einmal im Umlauf (gesetzlich oder nicht), so lässt sich dieser nur schwer stoppen.
- ▷ Die Daten, die bei den einzelnen Dienstanbietern liegen, sind nicht per se gefährlich für den Nutzer: Meist sind es nur kleine Mengen und für sich in keiner Weise relevant. Erst die Vernetzung der Daten ist das Problem, und diese geht zumeist nicht von nur einem Anbieter aus, so dass eine Regulierung hier nicht greifen kann.

3.2.2 Dienstanbieter

Die Dienstanbieter sind größtenteils nicht dazu verpflichtet, aktiv am Datenschutz für ihre Nutzer mitzuwirken, da ihre Dienste ja explizit dazu dienen, diese Daten zu veröffentlichen. Dies wird auch deutlich, wenn man die Datenschutzbestimmungen der führenden Anbieter betrachtet. So wird generell die Haftung für fehlerhaften Umgang mit den Daten abgewiesen, die Weitergabe an Dritte erlaubt etc.

Allerdings ließen sich durch die Anbieter viele Ideen für einen höheren Datenschutz umsetzten. Insbesondere ein graduelles Einstellen der Weitergabe und Nutzung von Daten durch andere Nutzer und Dienste würde schon viel ändern.

Auch kann im Marketing auf den besseren Datenschutz gegenüber Mitbewerbern hingewiesen werden, auch wenn für die Nutzer der Datenschutz leider nur ein Randkriterium bei der Entscheidung für einen Dienst darstellt.

3.2.3 Benutzer

Da weder der Gesetzgeber noch die Dienstanbieter viel zum Schutz der Daten beitragen können, muss der Anwender als *verantwortungsvoller Nutzer* selbst den größten Anteil beitragen. Er kann sich selbst für den Dienst entscheiden, der einerseits seine Wünsche erfüllt und andererseits die Datenschutzbedürfnisse befriedigt. Am wichtigsten ist aber, dass sich der Nutzer stets bewusst ist, dass er die Daten, die er auf allen diesen Plattformen zur Verfügung stellt, potenziell von der gesamten Welt gelesen werden können.

4 Schaden

Durch das geringe Bewusstsein der Nutzer, welche Gefahren von der Veröffentlichung von Informationen im *Web 2.0* ausgehen, muss diesen durch Beispiele klar gemacht werden, welche Konsequenzen dieses haben kann. Das einfachste Beispiel ist die Benutzung der Daten zu Marketingzwecken, also das Einbinden von personalisierter Werbung durch den Dienstanbieter oder aber die Weitergabe der Daten an andere Unternehmen. Leider ist hier das Bewusstsein der Bevölkerung nicht besonders ausgeprägt, vielmehr zeigt sich diese häufig zu solchen Maßnahmen bereit,

wie beispielsweise die Teilnahme an Gewinnspielen zeigt.

Genauso verhält es sich mit dem einfachen Abgreifen der Seiteninhalte durch Spiderprogramme: Die Nutzer glauben nicht, dass ihre Inhalten für irgendwen von Interesse sind. Wenn man sie allerdings auf die hohe Anzahl an Nutzerdaten aufmerksam macht, wird im allgemeinen doch eher erschrocken reagiert. So wurden die Daten eines der größten sozialen Netzwerke in Deutschland, StudiVZ, mehrfach[40] komplett extrahiert.

Im persönlichen Bereich gibt es unterschiedliche Dinge, die einzelne Nutzer betreffen können. Firmen haben wegen Einträgen in Weblogs, die sich negativ über die Firma äußerten, Mitarbeiter entlassen[41] – auch wenn die betroffenen Blogs nur kleine Leserzahlen hatten. Ein weitaus größeres Problem stellt Stalking dar, was einerseits virtuell, andererseits durchaus auch in der *normalen Welt* passieren kann. Häufigste Opfer hierbei sind Frauen[42] und Kinder[43].

Ein Bereich, in dem man relativ leicht die Nutzer überzeugen kann, dass die öffentliche Zugänglichkeit der Daten ein Problem darstellen könnte, ist die Bewerbung bei einem Arbeitgeber. Diese – oder eine von ihnen beauftragte Headhunting-Agentur – benutzen in zunehmendem Maße Suchmaschinen wie Google oder Technorati, um Informationen über die Bewerber zu bekommen.[44] Allerdings bereinigen viele Nutzer ihre Profile erst, wenn sie sich tatsächlich bewerben, auch wenn durchaus vorstellbar ist, dass gewisse Headhunting-Agenturen schon vorher die Datenbestände erfassen, um so ein aussagekräftigeres Profil des Bewerbers zu bekommen.

[40] http://studivz.irgendwo.org/,
http://turrican.unixag-zw.fh-kl.de/studianalyse/home,
http://www.buha.info/board/showpost.php?p=373844&postcount=35

[41] Beispielsweise Friendster: „Blog-Betreiber feuert Mitarbeiter wegen Blog-Eintrag“ (http://www.zdnet.de/news/tkomm/0,39023151,39125557,00.htm)

[42] Beispielsweise: „Neuer Ärger um StudiVZ: Sex-Stalker im Studentennetz“ (http://www.spiegel.de/netzwelt/web/0,1518,450866,00.html)

[43] Beispielsweise: “Four families have sued News Corp. and its MySpace social-networking site after their underage daughters were sexually abused by adults they met on the site.” (http://www.wtnh.com/Global/story.asp?S=5956321)

[44] In den USA ist dies schon ein weit verbreitetes Phänomen: “22.9 percent of the employers they surveyed are reviewing candidates’ profiles on social networking sites; 45.7 percent are using search engines like Google to do the same, and 14.3 percent review candidates’ personal Web sites and blogs.” (http://web.bsu.edu/careers/spotlight/articles/05-06/issue2/facebook.htm)

5 Fazit

Dieses Papier sollte zeigen, dass es kein Patentrezept zur Verbesserung des Datenschutzes im *Web 2.0*-Umfeld gibt. Die Gewalt liegt eindeutig beim Nutzer, im Guten wie im Schlechten. Er kann und muss entscheiden, welche Daten er bei welchen Diensten einstellen will und so der potenziell mitlesenden Weltbevölkerung zur Verfügung stellt.

6 Ausblick

Der Erfolg der sozialen Netzwerke und *Web 2.0*-Dienste ist offensichtlich. Die Nutzerzahlen steigen häufig exponentiell, so dass man sicher sein kann, dass derartige Angebote auch in der mittelfristigen Zukunft vorhanden sein werden und wahrscheinlich die Nutzung dieser auch noch ausweiten wird.

Die Informationen, die auf diesen Diensten bereitgestellt werden, sind sehr umfangreich und werden mehr und mehr auch den Alltag der Nutzer abbilden. Dadurch werden sehr private Details der jeweiligen Leben öffentlich. Auf der einen Seite wird noch gegen relativ geringe Informationsumfänge wie die Vorratsdatenspeicherung oder biometrische Daten in Pässen oder Melderegistern protestiert, während Millionen von Menschen viel weitreichender Daten in das Internet stellen. Auch wenn das eine staatlich angeordnet und das andere freiwillig erfolgt – die Gefahr, die davon ausgeht, sollte ernst genommen werden.