

# Datenschutz im Web 2.0 am Beispiel des sozialen Tagging-Systems BibSonomy

Beate Krause · Hana Lerch  
Andreas Hotho · Alexander Roßnagel  
Gerd Stumme

**Neben sozialen Netzwerken hinterlassen Nutzer auch in anderen typischen Web2.0-Anwendungen Spuren – zum Beispiel in sozialen Tagging-Systemen. Diese möglicherweise persönlichen Daten bieten eine interessante Grundlage für Data Mining und Profilbildung. Kann das Recht auf informationelle Selbstbestimmung dabei noch geschützt werden?**

## Einleitung

Soziale Tagging-Systeme (kooperative Verschlagwortungssysteme) gehören zu den in den vergangenen Jahren entstandenen Web2.0-Anwendungen. Die zumeist kostenlosen nutzbaren Angebote ermöglichen es ihren Benutzern, verschiedene Medien (z. B. Bookmarks, Fotos oder Videos) online abzuspeichern und

diese mit prägnanten, frei wählbaren Stichwörtern (Tags) zu beschreiben. Neben einer eigenständigen Strukturierung der eigenen Sammlung durch die Tags können Nutzer die öffentlichen Einträge anderer Anwender zur Informationsfindung nutzen: Dies wird durch eine einfache Navigationsstruktur ermöglicht, bei der nicht nur das Medium eines Eintrags, sondern auch die zugehörigen Tags und der einstellende Nutzer verlinkt sind. Durch die zentrale Speicherung hat der Benutzer außerdem jederzeit von jedem Rechner Zugriff auf seine Einträge.

Prominente Beispiele für soziale Tagging-Systeme sind Flickr (<http://www.flickr.com/>), welches die Verschlagwortung von Fotos ermöglicht, del.icio.us (<http://delicious.com/>), welches Lesezeichen verwaltet und CiteULike (<http://www.citeulike.org/>) oder Connotea (<http://www.connotea.org/>), welche sich auf die Verwaltung von Publikationsreferenzen konzentrieren. Unser eigenes System

BibSonomy dient der gemeinsamen Verwaltung von Lesezeichen und Publikationen (Kasten 1). Neben einer Vielfalt von unterschiedlichen Medien, die in den Systemen verwaltet werden können, unterscheiden sich die existierenden Systeme durch die angebotenen Funktionalitäten (z. B. Exportfunktionen, APIs) und den Grad der sozialen Interaktion (anonyme Nutzung bis Anzeige eines Benutzerprofils). Eine Gemeinsamkeit aller Systeme ist das zugrunde liegende Datenmodell, das mit dem Kunstwort „Folksonomy“ beschrieben wird. Der Begriff setzt sich aus „folk“ und „taxonomies“ zusammen und bezeichnet eine von Nichtexperten kreierte Kategorisierung von Inhalten. Hierbei entstehen Strukturen, indem viele Nutzer die gleichen Ressourcen mit verschiedenen beschreibenden Tags versehen.

Die zunehmende Nutzung sozialer Tagging-Systeme hat zu einem rasanten Wachstum von frei zugänglichen Daten geführt. Neben den auch in sozialen Netzwerken anfallenden Profildaten werden in Tagging-Systemen zusätzlich die von Nutzern eingestellten Einträge abgespeichert. Auch dabei hinterlassen Nutzer Spuren, die Ausdruck ihrer eigenen Interessen, Einstellungen und menschlichen Beziehungen werden können. Aufgrund der öffentlichen Speicherung der Einträge sind diese zumeist beliebig im Internet abrufbar.

DOI 10.1007/s00287-010-0485-8  
© Springer-Verlag 2010

Beate Krause · Andreas Hotho · Gerd Stumme  
Fachgebiet Wissensverarbeitung,  
Universität Kassel, Kassel  
E-Mail: {krause, hotho, stumme}@cs.uni-kassel.de

Hana Lerch · Alexander Roßnagel  
Projektgruppe Verfassungsverträgliche Technikgestaltung,  
Universität Kassel, Kassel  
E-Mail: {hana.lerch, a.rossnagel}@uni-kassel.de

## Zusammenfassung

Soziale Tagging-Systeme gehören zu den in den vergangenen Jahren entstandenen Web2.0-Systemen. Sie ermöglichen es Anwendern, beliebige Informationen in das Internet einzustellen und untereinander auszutauschen. Je nach Anbieter verlinken Nutzer Videos, Fotos oder Webseiten und beschreiben die eingestellten Medien mit entsprechenden Schlagwörtern (Tags). Die damit einhergehende freiwillige Preisgabe oftmals persönlicher Informationen wirft Fragen im Bereich der informationellen Selbstbestimmung auf. Dieses Grundrecht gewährleistet dem Einzelnen, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen. Für viele Funktionalitäten, wie beispielsweise Empfehlungsdienste oder die Bereitstellung einer API, ist eine solche Kontrolle allerdings schwierig zu gestalten. Oftmals existieren keine Richtlinien, inwieweit Dienstanbieter und weitere Dritte diese öffentlichen Daten (und weitere Daten, die bei der Nutzung des Systems anfallen) nutzen dürfen. Dieser Artikel diskutiert anhand eines konkreten Systems typische, für den Datenschutz relevante Funktionalitäten und gibt Handlungsanweisungen für eine datenschutzkonforme technische Gestaltung.

Schon heute wird diskutiert, z. B. in [6], inwieweit sich diese Daten für eine Verbesserung von Suchergebnissen, personalisierter Werbung oder der Profilbildung von Internetnutzern (z. B. für potenzielle Arbeitgeber) nutzen lassen. Oftmals bleibt dabei eine Betrachtung der zu schützenden Bereiche von Nutzern, Betreibern und häufig auch Dritten erst einmal im Hintergrund. Die Entwicklung, der Betrieb und die Nutzung von sozialen Tagging-Systemen berühren aber die Rechtssphären dieser Personen und Institutionen. So stellen sich angesichts der anfallenden Daten, die auch in Beziehung zu einzelnen Nutzern stehen, insbesondere Fragen zum Datenschutzrecht, aber auch dem Urheberrecht, dem Wettbewerbsrecht, dem Jugendschutzrecht oder dem Strafrecht.

In vielen Fällen fehlt ein fachlicher Austausch, bei dem Informatiker und Juristen gemeinsam die Möglichkeiten und Grenzen der neuartigen

Anwendungen betrachten und diese entsprechend gestalten. Ein erster Schritt ist die Analyse in [5]. Im Folgenden soll eine technisch-rechtliche Auseinandersetzung mithilfe einer repräsentativen Anwendung erfolgen. Bei dieser handelt es sich um BibSonomy, ein von einem Teil der Autoren entwickeltes und betriebenes soziales Tagging-System für Lesezeichen und Publikationen (Kasten 1). Im Folgenden werden vom Nutzungsbeginn über die verschiedenen Anwendungsmöglichkeiten hinweg bis zur Beendigung der Mitgliedschaft Fragen des Datenschutzrechts untersucht und Möglichkeiten für eine auf den Schutz der Betroffenen abgestimmte Gestaltung vorgestellt. Die Studie wurde im Rah-

Das soziale Tagging-System BibSonomy [1, 4] ist seit Ende Dezember 2005 unter der Adresse <http://www.bibsonomy.org/> online und wird vom Fachgebiet Wissensverarbeitung der Universität Kassel betrieben. Als eine Art wissenschaftliches Bücherregal im Internet ermöglicht BibSonomy die Onlineverwaltung von Lesezeichen und Publikationen. Die primäre Zielgruppe des Systems sind universitäre Anwender (Studenten, Wissenschaftler, Forschungsgruppen). Deswegen werden spezielle Funktionalitäten angeboten, die Nutzer bei der systematischen Kategorisierung von Publikationen, dem Literaturtausch und ihrer Recherche unterstützen. BibSonomy ist auch die Basis für die in dem DFG-Projekt „PUMA – Akademisches Publikationsmanagement“ entwickelte universitäre Publikationsberichtsplattform PUMA (<http://www.puma.uni-kassel.de/>).

Das Fachgebiet nutzt BibSonomy für die Evaluierung von Algorithmen und Methoden des Information Retrieval und der Wissensentdeckung. In der Ansammlung der Nutzerdaten können interessante Strukturen entdeckt und diese algorithmisch aufbereitet werden. Die entwickelten Verfahren werden wiederum im System implementiert und können dadurch zu einer verbesserten Nutzerfreundlichkeit beitragen. Eine detaillierte Übersicht über Forschung und ihre Anwendungen in BibSonomy findet man in [1].

### Kasten 1 BibSonomy

## Abstract

Social tagging systems form part of the recently developed Web 2.0 systems, which enable users to publish and share user-defined material in the web. Depending on the service provider, users can link videos, photos or web sites and describe them with arbitrarily selected tags. The resulting voluntary disclosure of often private information in the internet raises questions concerning the right to control this data. For many of the typical features of social tagging systems (for example recommender systems or the provision of an API) such control mechanisms are difficult to implement. Often, there exist no guidelines to what extent providers and other third parties can use the available data. This article discusses such relevant functions on the basis of an existing social tagging system and provides guidelines for a privacy-preserving technical design.

men des DFG-Projekts „Info 2.0 – Informationelle Selbstbestimmung im Web 2.0“ durchgeführt.

## Rechtliche Grundlagen

Ausgangspunkt der rechtlichen Bewertung einzelner Datenverarbeitungsvorgänge stellt das Recht auf informationelle Selbstbestimmung als Ausprägung des allgemeinen Persönlichkeitsrechts nach Art. 2 Abs. 1 und Art. 1 Abs. 1 des Grundgesetzes dar. Das Bundesverfassungsgericht hat dieses Recht in seinem Volkszählungsurteil aus dem Jahr 1983 konkretisiert und als die Befugnis des Einzelnen beschrieben, grundsätzlich selbst über die Preisgabe und Verwendung seiner persönlichen Daten zu bestimmen [2]. Geschützt sind demnach nicht die Daten selbst, sondern der Inhaber der Daten vor unbefugter Verwendung seiner Daten durch den Staat, aber auch durch private Stellen. Für den Bereich der sozialen Tagging-Systeme relevante einfachgesetzliche Regelungen finden sich im Bundesdatenschutzgesetz (BDSG) und im Telemediengesetz (TMG).

Das BDSG unterscheidet beim Umgang mit Daten die Phasen der Datenerhebung, -verarbeitung und -nutzung, wobei die Rechtmäßigkeit für jeden einzelnen Vorgang gesondert zu prüfen ist. Der Umgang mit personenbezogenen Daten ist nach § 4 Abs. 1 BDSG, für Bestands- und Nutzungsdaten

nach § 12 Abs. 1 TMG nur zulässig, wenn es das Gesetz erlaubt oder der Betroffene einwilligt. Für einen Zweck erhobene Daten dürfen nur für diesen Zweck verarbeitet und genutzt werden, es sei denn eine Norm oder die Einwilligung des Betroffenen gestatten die Zweckänderung. Eine Datenspeicherung auf Vorrat ist damit grundsätzlich unzulässig, sofern nicht besondere Bestimmungen greifen. Es sind so wenig personenbezogene Daten wie möglich zu erheben und zu verwenden und so bald wie möglich zu löschen oder zu anonymisieren. Für die Gestaltung eines Systems bedeutet dies, dass von vornherein die technische Variante zur Erreichung eines Ziels zu wählen ist, die am wenigsten Daten erfordert. Dabei soll der Einzelne jedoch nicht vor jeglicher Verwendung seiner Daten geschützt werden, sondern selbstbestimmt darüber entscheiden können, welche Daten er verarbeiten lassen möchte. Daher hat der Umgang mit den Daten transparent zu erfolgen, damit der Betroffene seine Rechte wahrnehmen kann. Dazu sieht das Gesetz verschiedene Informationspflichten des Datenverwenders und Auskunftsrechte des Betroffenen vor.

Es ist zwischen *Bestands-* und *Nutzungsdaten* zu unterscheiden, die unter die Datenschutzregeln des TMG fallen, und *Inhaltsdaten*, die nach BDSG zu beurteilen sind. Bestandsdaten sind solche personenbezogenen Daten, die vom Anbieter eines Telemediendienstes – wie etwa einem sozialen Tagging-System – z. B. bei der Registrierung erhoben und für die Begründung, die inhaltliche Ausgestaltung oder die Änderung des Nutzungsverhältnisses verwendet werden. Das können der reale Name, ein Nutzernamen, die E-Mailadresse oder ein Passwort sein, aber auch jegliche Arten von Informationen, die zu einer Personalisierung des Angebotes genutzt werden, wenn dies dem vom Nutzer gewählten Dienst entspricht. Nutzungsdaten sind diejenigen personenbezogenen Daten des Nutzers, die dem Nutzer die Nachfrage nach dem Telemedium ermöglichen und während der Nutzung des Dienstes entstehen, z. B. durch Interaktionen des Nutzers mit dem Diensteanbieter. Dazu gehören z. B. die IP-Adresse des Rechners des Nutzers, die Zeit und Dauer der Nutzung und die abgerufenen Inhalte, aber auch gegebenenfalls der Nutzernamen – der zugleich ein Bestandsdatum darstellt. Inhaltsdaten sind diejenigen Daten, die der Nutzer im Rahmen seiner Interaktion mit dem Anbieter in Eingabefelder eingibt und die in keinem

funktionalen Zusammenhang mit der Nutzung des Dienstes als Telemedium stehen. Dazu zählen Eingaben zu Lesezeichen, hochgeladene Inhalte, Einträge zur eigenen Person oder Kritiken zu anderen Nutzern, unabhängig davon, ob diese Einträge für andere Nutzer öffentlich sind oder nicht. Die Zulässigkeit der Erhebung und Verwendung von Inhaltsdaten richtet sich nach dem BDSG. Einzelne Daten können je nach Verwendungszusammenhang auch mehreren Kategorien zuzuordnen sein. Die Zulässigkeit von deren Erhebung und Verwendung richtet sich dementsprechend nach der für die jeweilige Verwendung geltenden Regelung.

Die Vorschriften finden mit wenigen Ausnahmen nur dann Anwendung, wenn es sich bei den verwendeten Daten um personenbezogene Daten handelt. Das sind nach § 3 Abs. 1 BDSG Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbarer natürlicher Person. Dazu zählen alle Arten von Informationen, die in Beziehung zu einer Person stehen können, wie z. B. Name, Alter, Geschlecht, Interessen oder Bekanntschaften. Bestimmt ist die Person, wenn die Daten selbst einen unmittelbaren Rückschluss auf die Identität der Person ermöglichen, wie es etwa beim Namen oder einer nur auf eine Person zutreffenden Anschrift der Fall ist. Bestimmbar ist die Person, wenn sich ihre Identität zwar nicht aus den Daten selbst, aber mithilfe weiterer Informationen feststellen lässt, die der jeweiligen datenverwendenden Stelle zur Verfügung stehen. Dieselbe Information kann demnach für den einen Datenverwender ein personenbezogenes Datum darstellen, für einen anderen ein nichtpersonenbezogenes [8]. Der Umgang mit nichtpersonenbezogenen Daten unterliegt grundsätzlich nicht dem Datenschutzrecht.

Die Personenbezogenheit kann sich auch aus der Kombinationsmöglichkeit der Datenarten ergeben, sodass z. B. die Hinterlegung eines personenbezogenen Datums wie dem realen Namen oder einer den realen Namen enthaltenden E-Mailadresse bei den Bestandsdaten nicht nur die restlichen Bestandsdaten zu personenbezogenen macht. Über die Verknüpfung mit dem – dann ebenfalls personenbezogenen – Nutzernamen erhalten auch die Nutzungs- und Inhaltsdaten einen Personenbezug, sodass auch deren Verwendung auf ihre datenschutzrechtliche Zulässigkeit nach TMG und BDSG zu prüfen ist. Dies gilt umgekehrt auch für ein unter

einem Pseudonym geführtes Nutzerkonto, dessen sämtliche Daten personenbezogen werden, sobald ein Eintrag einen Hinweis auf die hinter dem Nutzer stehende reale Person gibt [9], etwa durch namentliche Nennung des Nutzers als Autor eines Textes. Da der Anbieter aufgrund der automatisierten Vorgänge und der Menge an Daten und Kombinationsmöglichkeiten nicht im Einzelnen feststellen kann, bei welchen Daten für ihn ein Personenbezug tatsächlich herstellbar ist und bei welchen nicht, hat er, sofern er nicht durch die entsprechende Systemgestaltung sicherstellt, dass Daten ausschließlich ohne Möglichkeit der Personenbeziehbarkeit erhoben und verwendet werden, sämtliche Daten vorsorglich als personenbezogene zu behandeln und bei deren Verwendung die Vorschriften von TMG und BDSG, insbesondere die o. g. Grundsätze, zu beachten.

BibSonomy stellt insofern eine Besonderheit dar, als dass die Betreiberin eine hessische Hochschule und damit eine unter Landesaufsicht stehende juristische Person ist. Somit unterfällt die datenschutzrechtliche Bewertung des Umgangs mit Daten in diesem speziellen Fall dem hessischen Datenschutzgesetz (HDSG). Wegen der hohen Praxisbedeutung werden im Folgenden jedoch die typischen Funktionen beim Social Bookmarking anhand des Modells von BibSonomy rechtlich so beurteilt, als handle es sich um einen privaten Anbieter, und die Regelungen des BDSG herangezogen. Im Einzelfall können sich für BibSonomy davon abweichende Bewertungen ergeben, auf die am Ende des Textes eingegangen wird. Ein grundsätzlicher Unterschied liegt darin, dass im Rahmen von Forschungsvorhaben ein weiterer Handlungsspielraum hinsichtlich der Daten- und Methodenauswahl besteht, um in diesem Bereich Erkenntnisse gewinnen zu können. Private Anbieter sind hingegen nicht frei in der Wahl ihrer Mittel, sondern haben auf die vorhandenen wissenschaftlich gewonnenen Erkenntnisse zurückzugreifen.

Beginnend bei der Aufnahme von Daten im Rahmen der Registrierung (Bestandsdaten), der Speicherung von Objekten wie Lesezeichen, Fotos oder selbst gewählten Tags (Inhaltsdaten) bis zur Protokollierung der Nutzung des Dienstes (Nutzungsdaten) soll im Folgenden am Beispiel BibSonomy aufgezeigt werden, an welchen Stellen personenbezogene Daten in sozialen Tagging-Systemen anfallen können und wie deren Verwendung im Einzelnen rechtlich zu bewerten ist. Für die

datenschutzgerechte Gestaltung solcher Systeme ist es erforderlich, etwaige Gefahren zu erkennen und wo möglich technisch zu minimieren.

## Registrierung

Der erste Schritt zur Verwaltung der eigenen Lesezeichen und Publikationsdaten in BibSonomy ist das Anlegen eines Benutzerkontos. Für die Registrierung werden der Benutzername, der wirkliche Name, die E-Mailadresse, die Homepage und ein Passwort (sowie ein Captcha) abgefragt. Dem Nutzer wird angezeigt, welche Informationen notwendig (Benutzername, E-Mailadresse, Passwort), und welche optional für die Anmeldung sind. Zusätzlich müssen Nutzer den AGBs und Datenschutzbestimmungen zustimmen, wenn sie sich anmelden möchten. In den eigenen Benutzereinstellungen können Anwender über die Sichtbarkeit ihrer Daten verfügen, Informationen löschen, ändern oder hinzufügen. So können Nutzer etwa Daten wie den wirklichen Namen oder die Homepage im Rahmen eines Lebenslaufes oder eines Profils für andere Nutzer sichtbar machen.

Nach § 12 Abs. 1 TMG darf der Anbieter personenbezogene Daten zur Bereitstellung von Telemedien nur erheben und verwenden, soweit eine auf Telemedien bezogene Norm es erlaubt oder der Nutzer eingewilligt hat. Die Tatsache, dass ein Nutzer freiwillig Angaben zu seiner Person, z. B. in ein Registrierungsformular, eingibt, stellt dabei noch keine Einwilligung in diesem Sinne dar. Ohne wirksame Einwilligung des Nutzers (s. Kasten 2) richtet sich die Zulässigkeit der Datenverwendung nach § 14 Abs. 1 TMG. Als Bestandsdaten des Nutzers dürfen nur diejenigen personenbezogenen Daten zulässigerweise erhoben werden, die für die Begründung, inhaltliche Ausgestaltung oder Änderung eines Vertragsverhältnisses zwischen dem Diensteanbieter und dem Nutzer über die Nutzung von Telemedien erforderlich sind. Sie dürfen auch nur zu diesem Zweck verwendet werden.

Ob die Daten für das Nutzungsverhältnis erforderlich sind, richtet sich nach dessen Inhalt. So kann bei kostenpflichtigen Diensten die Angabe des Realnamens und einer Kontoverbindung erforderlich sein. Bietet der Dienst als zentrales Merkmal eine auf Einzelbedürfnisse des Nutzers zugeschnittene Nutzung, können alle für die Personalisierung erforderlichen Angaben zulässigerweise erhoben und für diesen Zweck verwendet werden. Zu einem

Im Bereich der Telemedien, unter die auch die sozialen Tagging-Systeme fallen, kann die Einwilligung des Nutzers elektronisch erklärt werden, z. B. durch das Setzen eines Hakens in einer Checkbox neben einer vorformulierten Einwilligungserklärung. Ebenfalls ausreichend ist die bestätigende Wiederholung des Übermittlungsbefehls durch den Nutzer, wenn dabei zumindest Ausschnitte der Einwilligungserklärung auf dem Bildschirm angezeigt werden. Jedoch hat der Anbieter für diese Art der Einwilligung die Voraussetzungen des § 13 Abs. 2 Nr. 1 bis 4 TMG sicherzustellen: Er muss technisch dafür Sorge tragen, dass der Nutzer die Einwilligung bewusst und eindeutig in Bezug auf den jeweiligen Verwendungszweck und die betroffenen Daten erteilen kann. Sie muss protokolliert werden und für den Nutzer jederzeit abrufbar sein. Der Anbieter muss sicherstellen, dass der Nutzer jederzeit die Möglichkeit hat, seine Einwilligung zu widerrufen. Erfüllt der Anbieter die genannten Voraussetzungen, so ist dennoch zu beachten, dass die Datenverwendung zu anderen Zwecken als den für den Nutzer erkennbaren nicht durch die Einwilligung gedeckt ist.

### Kasten 2 Einwilligungserklärung

anderen Zweck ist die Verwendung wiederum nur zulässig, wenn eine Einwilligung des Nutzers vorliegt. Bietet ein System keine Personalisierung an, für die es die bei der Registrierung erhobenen Daten verwendet, ist grundsätzlich nur die Einholung von für die Einrichtung eines passwortgeschützten Nutzerkontos unerlässlichen Informationen wie Nutzernamen und Passwort zulässig. Die Erhebung und entsprechende Verwendung des Realnamens beispielsweise für die vereinbarungsgemäße Veröffentlichung eines Lebenslaufs ist zulässig. Gibt es keine Funktion, die dessen Verwendung erfordert, ist die Erhebung unzulässig. Gehört es zum vereinbarten Nutzungsinhalt, per E-Mail z. B. über Neuigkeiten im System, zu technischen Störungen oder mit Informationen zum eigenen Konto benachrichtigt zu werden, so ist die Erhebung und Verwendung der E-Mailadresse zu diesem Zweck zulässig, ansonsten und zu anderen Zwecken jedoch nicht.

Der Anbieter muss nicht hinnehmen, dass das von ihm betriebene System durch den Missbrauch von Spammern behindert oder gar in seinem Bestand gefährdet wird. Insofern ist ihm ein „Selbstverteidigungsrecht“ zuzugestehen, welches die geeigneten Maßnahmen zur Identifizierung der ungewollten Kundschaft rechtfertigt. Jedoch sind dem im Interesse der legitimen Nutzer und ihrer Rechte Grenzen gesetzt. Ob ein Bestandsdatum für die Identifizierung von Spammern erhoben und herangezogen werden darf, hängt von dessen Verlässlichkeit als Indikator ab. Je zuverlässiger ein Kriterium reguläre von nichtregulären Nutzern unterscheiden lässt, desto eher wird dessen Erhebung und Verwendung zu diesem Zweck als erforderlich und damit zulässig anzusehen sein. Insofern sind Anbieter jedoch, soweit vorhanden, auf die einschlägigen Erkenntnisse der Forschung zu verweisen.

### **Kasten 3 Spambekämpfung**

Ein weiterer Zweck für die Erhebung und Verwendung von Bestandsdaten besteht in der Bekämpfung von Spam (s. Kasten 3), denn ein Großteil neuer BibSonomy-Nutzer sind Spammer, die massenhaft Werbung oder nicht jugendfreie Inhalte in das System einstellen. Diese Nutzer sollen möglichst früh erkannt und aus allen öffentlich sichtbaren Seiten entfernt werden. Dafür wird ein Spamfilter eingesetzt, der die Nutzer mit bekannten Klassifikationsverfahren in vier Gruppen aufteilt: Spammer und Nichtspammer, jeweils aufgeteilt als solche, die sicher bzw. nur mit weniger hoher Sicherheit als solche klassifiziert werden können. Als Merkmale werden Informationen aus den bei der Registrierung eingegebenen Daten sowie den eingestellten Einträgen verwendet. Nutzer, die das System zum Einstellen von Spam verwenden möchten, wählen beispielsweise oft E-Mailadressen oder Nutzernamen mit Ziffern. Außerdem sind Konten, bei denen ein voller realer Name (Vor- und Nachname) angegeben wird, zumeist keine Spamkonten. Oft fügen Spammer ihren Einträgen viele Tags hinzu, damit die Einträge leicht auffindbar werden. Außerdem nutzen Spammer ein ähnliches Vokabular und verschlagworten mitunter die gleichen Ressourcen. Diese Merkmale lassen sich gut

für die Entdeckung von Spam einsetzen – insbesondere wenn der Nutzer erst wenige Einträge im System gespeichert hat. Zudem wird angenommen, dass Nutzer mit einer universitären E-Mailadresse legitime Nutzer sind und nicht weiter untersucht werden müssen.

*Handlungsanweisung:* Entsprechend dem Grundsatz der Datensparsamkeit sind generell so wenige Daten wie möglich zu erheben und zu verwenden. Im Einzelfall ist zu überlegen, welche Daten für die Durchführung des angebotenen und vom Nutzer so gewählten Dienstes tatsächlich erforderlich sind. Nur diese dürfen erhoben und verwendet werden, und auch nur für diesen Zweck. Dabei kann ein personalisierter Dienst mehr Angaben erfordern als ein nichtpersonalisierter. Nicht erforderliche Daten dürfen nur erhoben und verwendet werden, wenn und soweit eine Einwilligung des Nutzers vorliegt (s. Kasten 2).

### **Einstellen von Inhalten**

Eine der ersten Interaktionen mit BibSonomy ist das Abspeichern einer Ressource, also eines Weblesezeichens oder der Metadaten einer Publikation. Im ersten Fall übernimmt der Benutzer den Link zu einer Webseite und fügt Tags als Beschreibung hinzu. Um Publikationsdaten im System zu speichern, stehen mehrere Möglichkeiten zur Verfügung, beispielsweise die automatische Extraktion von Daten aus einer auf einer Webseite markierten Referenz oder die Verwendung von Scrapern bei digitalen Bibliotheken. Zusammen mit den Metadaten (z. B. Autor, Titel, Zeitschrift, DOI) lassen sich zusätzlich noch Tags, Beschreibungen und Kommentare abspeichern.

Entsprechend dem sich aus dem Nutzungsverhältnis ergebenden Zweck speichert der Anbieter diese Daten, um sie für den Nutzer bereitzuhalten und gegebenenfalls im Internet zu veröffentlichen. Da diese Vorgänge Hauptzweck des Nutzungsverhältnisses sind und die Erhebung, Speicherung und Übermittlung über das Web zur Erreichung dieses Zwecks erforderlich sind, darf der Anbieter die Daten gemäß § 28 Abs. 1 S. 1 Nr. 1 BDSG zunächst einmal zu genau diesem Zweck und in dieser Weise verwenden. Auch die Analyse von Inhaltsdaten, um Nutzer zu identifizieren, die das System zur Verbreitung von Spameinträgen missbrauchen, ist zulässig. Indem sie einen spamfreien Raum erhält, dient diese Art der Inhaltsdatenverarbeitung

nicht nur der Attraktivität, sondern auch der sinnvollen Nutzbarkeit des Angebotes für ordentliche Nutzer und bewahrt zugleich die ihnen zugedachten Speicherkapazitäten. Daher ist diese Verarbeitung notwendig und ebenfalls zulässig.

Bei der Verwendung von Inhaltsdaten in Data-Mining-Verfahren ist zunächst festzustellen, ob dies der Erfüllung vertraglicher Zwecke dient, also notwendig ist, um vereinbarte Funktionen ausführen zu können oder gar selbst eine Funktion darstellt. In diesem Fall ist die entsprechende Verwendung zulässig. Dies kann je nach Ausgestaltung des Nutzungsverhältnisses bei dem Einsatz von Algorithmen zur Empfehlung von Tags, Ressourcen oder auch anderen Nutzern der Fall sein. Dass die Funktion im Interesse des Nutzers liegt, ist dabei nicht allein entscheidendes Kriterium, der Nutzer muss sich ihrer bei der Wahl des Angebotes auch bewusst gewesen sein. Zu anderen Zwecken kann die Verwendung zwar grundsätzlich im Einzelfall nach § 28 Abs. 1 S. 1 Nr. 2 und Nr. 3 BDSG zulässig sein, wenn dem kein überwiegendes schutzwürdiges Interesse des Nutzers entgegensteht. Da diese Abwägung jedoch nur im Einzelfall und nicht pauschal vom Anbieter durchzuführen ist, scheidet eine Rechtfertigung nach diesen Vorschriften regelmäßig bei automatisierten Vorgängen, wie sie für den Bereich der Onlineanwendungen typisch sind, aus. Die Profilbildung aus Inhaltsdaten zum Zwecke der Schaltung von personalisierter Werbung setzt nach § 28 Abs. 3 BDSG im Regelfall die Einwilligung des Nutzers voraus und darf ohne eine solche nicht vorgenommen werden. Für die Einwilligung gilt das oben Gesagte.

*Handlungsanweisung:* Sinnvolle und für Bookmarking-Systeme typische Funktionen, zu deren Durchführung die Verwendung der eingestellten Inhaltsdaten erforderlich ist, sollten bei Begründung des Nutzungsverhältnisses möglichst klar zu dessen Inhalt gemacht werden, etwa durch eine katalogartige Aufzählung bei der Registrierung. Andernfalls kann es zu Unklarheiten kommen und einzelne Funktionen zu Lasten des Anbieters als nicht zum Vertragsinhalt gehörend eingestuft werden, was den dazu erforderlichen Umgang mit personenbezogenen Inhaltsdaten ohne Einwilligung des Nutzers unzulässig machen kann. Dies gilt unabhängig davon, dass im Rahmen der Datenschutzerklärung die entsprechenden Datenverwendungen ebenfalls zu erläutern sind.

### **Verwaltung von Publikationsmetadaten**

Gelegentlich beschwerten sich Dritte über Einträge anderer Nutzer, in denen sie (beispielsweise als Autoren) genannt werden. Die Beschwerden richten sich zumeist per E-Mail an die Betreiber. So kann der Autorenname falsch geschrieben sein, ein Autor kann nicht oder fälschlicherweise genannt sein oder die Zeitschrift/Konferenz wurde fehlerhaft eingetragen. Die E-Mailabsender fordern den Betreiber zumeist auf, die ihrer Angabe nach nicht korrekten Metadaten zu berichtigen. Obwohl technisch möglich, würde es gegen die Natur der eigenständigen Verwaltung in sozialen Tagging-Systemen sprechen, wenn Einträge nicht von ihren Nutzern, sondern vom Anbieter geändert würden.

Metadaten sind einerseits – gegebenenfalls personenbezogene – Inhaltsdaten des Nutzers. Davon unabhängig stellen sie zugleich jedoch auch Einzelangaben zu persönlichen oder sachlichen Verhältnissen der jeweils von ihnen beschriebenen Autoren, unter Umständen auch weiterer Personen wie Verleger oder Mitautoren, und somit deren personenbezogene Daten, dar, sofern sie anhand der Metadaten selbst oder anhand von Zusatzwissen identifizierbar sind. Eine solche Identifizierung ist bei Publikationen häufig durch die Zuhilfenahme von Suchmaschinen möglich.

Nach § 35 Abs. 1 BDSG sind unrichtige personenbezogene Daten zu berichtigen. Sind Metadaten falsch, sind sie demnach von der verantwortlichen Stelle richtigzustellen. Wer hinsichtlich der vom Nutzer eingegebenen Daten verantwortliche Stelle ist, also nach § 3 Abs. 7 BDSG die Person oder Stelle, die personenbezogene Daten für sich selbst erhebt, verarbeitet oder nutzt oder dies durch andere im Auftrag vornehmen lässt, ist in der Konstellation des „user generated content“ nicht abschließend geklärt. Vieles spricht jedoch dafür, sowohl den einstellenden Nutzer als auch den Systembetreiber als verantwortlich anzusehen. Für die Verantwortlichkeit des einstellenden Nutzers spricht, dass er die Daten aus eigenem Antrieb auswählt und eingibt und den Betreiber mittels seiner Eingaben auffordert, diese Angaben zu speichern und Dritten zur Verfügung zu stellen. Für den Betreiber gilt hingegen, dass dieser allein technisch betrachtet direkten Einfluss auf die Daten besitzt und mittels seiner technischen Infrastruktur den Abruf der Inhalte durch andere Nutzer ermöglicht. Der Betreiber kann sich demnach nicht auf die alleinige Verantwortlichkeit

des Nutzers berufen, sondern hat die ihm bekannten falschen Metadaten zu berichtigen.

Der Betreiber muss keinen unverhältnismäßigen Aufwand zur Erforschung der Wahrheit betreiben, häufig wird bereits eine Suchmaschinenanfrage weiterhelfen können. Den Nachweis der Betroffenheit kann der jeweilige Autor z. B. durch Vorlage eines Identitätsnachweises, den der inhaltlichen Richtigkeit seines Anliegens mit bibliografischen Mitteln liefern. Verbleiben dem Betreiber trotz einer mit den ihm zur Verfügung stehenden Mitteln durchgeführten Überprüfung Zweifel am Wahrheitsgehalt etwaiger Beanstandungen, hat er die Daten gegen den weiteren Abruf zu sperren, bis eine Klärung herbeigeführt werden kann.

*Handlungsanweisung:* Da die datenschutzrechtliche Verantwortlichkeit von Betreibern nicht abschließend geklärt ist, sollte der Betreiber schon im Eigeninteresse dafür sorgen, dass unrichtige Metadaten berichtigt werden. Erhält er Kenntnis von deren Unrichtigkeit, kann er in einem ersten Schritt dem Nutzer Gelegenheit geben, die Korrektur selbst vorzunehmen. Reagiert dieser nicht oder weigert sich, sollte der Betreiber den entsprechenden Eintrag sperren oder selbst ändern. Wenn unklar ist, ob eine Beanstandung berechtigt ist, sollte der Betreiber, solange kein unverhältnismäßiger Aufwand betrieben werden muss, die Angelegenheit prüfen. Kann die Frage nicht geklärt werden, so sollte er die Daten sperren.

## **Nutzung des Systems zur Informationssuche**

Ähnlich zu Suchmaschinen unterstützen soziale Tagging-Systeme Nutzer bei der Entdeckung von interessanten Informationen. Dies wird durch eine spezifische Navigationsstruktur unterstützt, die dem Nutzer ein einfaches Browsing ermöglicht. Dabei werden die zu einem Eintrag gehörenden Tags sowie der Nutzer, der den Eintrag eingestellt hat, verlinkt. Ausgehend von ihren eigenen Einträgen oder Einträgen allgemeiner Übersichtsseiten (zum Beispiel der populärsten oder der aktuellsten Einträge) können Nutzer auf Tags oder auf die Namen anderer Nutzer klicken und so Einträge mit den gleichen oder ähnlichen Tags sowie Einträge, die von einem spezifischen Nutzer eingestellt wurden, sehen. In [3] wurden die Small-World-Eigenschaften für Folksonomien nachgewiesen. Dies bedeutet, dass ein Nutzer in nur wenigen Schritten von einem the-

matischen Bereich in einen völlig anderen gelangen kann. In BibSonomy sind dies durchschnittlich drei Navigationsschritte. Zusätzlich sind stark vernetzte Knoten dieses Linkgraphen fast immer auch thematisch eng verwandt. So können Nutzer beim Stöbern in ihrer nächsten Nachbarschaft öfters einmal über interessante Informationen „stolpern“.

Neben einer einfachen Navigationsstruktur bieten soziale Tagging-Systeme eine Suchfunktion an, mit der gezielt nach Einträgen zu einem bestimmten Thema gesucht werden kann. Oft werden die Ergebnisse sortiert, sei es nach der Häufigkeit der eingestellten Einträge oder nach den spezifischen Interessen des Anfragenden.

Bei der Nutzung des Systems zur Informationsfindung zeigen Nutzer durch die Eingabe von Suchbegriffen, durch Anklicken von Hyperlinks oder durch das Kopieren von öffentlichen Einträgen Dritter ihre Interessen an bestimmten Inhalten. Dieses Verhalten lässt sich u. a. durch Angaben im Protokoll des Webservers (z. B. zu IP-Adresse, Eingangsdatum oder Referrer), durch Javaskript-Techniken zur Verfolgung der Nutzerinteraktion mit der Seite und durch das Setzen einer eindeutig bestimmten Identifikationsnummer in Form eines Cookies mitverfolgen. Besonders ist dabei, dass diese als Nutzungsdaten einzuordnenden Informationen nicht nur einer einzelnen Cookie-ID, sondern bei registrierten und eingeloggtten Nutzern ebenfalls dem jeweiligen Nutzernamen zugeordnet werden können. Auf diese Weise lassen sich ganze Nutzungsprofile Einzelner bilden. Diese Nutzungsprofile lassen sich für eine Verbesserung des Angebotes (zum Beispiel ein verbessertes Ranking) verwenden, sie könnten auch an Werbeanbieter weitergegeben werden, die anhand der Profile personalisierte Werbung schalten können. Dies ist allerdings in BibSonomy nicht der Fall.

Nutzungsdaten und -profile sind für den Betreiber personenbezogene Daten des Nutzers, wenn der Betreiber über die Möglichkeiten verfügt, sie der hinter dem Nutzer stehenden Person zuzuordnen. Diese Möglichkeit kann sich zum einen aus den bei der Registrierung hinterlegten Daten ergeben, wenn sich aus diesen wiederum die Identität des Nutzers ergibt, was beispielsweise bei Hinterlegung des Klarnamens oder einer den vollen Namen wiedergebenden E-Mailadresse der Fall sein kann. Aber auch aus den vom Nutzer eingegebenen Inhalten können sich solche Zuordnungsmöglichkeiten ergeben.



Sofern der Betreiber demnach keine technischen Maßnahmen ergreift, die Zuordenbarkeit der Nutzungsdaten zu einem bestimmten Nutzerkonto und den damit verbundenen Erkenntnisquellen auszuschließen, sind sämtliche Nutzungsdaten vom Betreiber als personenbezogen zu behandeln, da es für ihn nicht feststellbar ist, ob der Personenbezug im Einzelfall hergestellt werden kann oder nicht.

Nach § 15 Abs. 1 S. 1 TMG dürfen personenbezogene Nutzungsdaten ohne ausdrückliche Einwilligung des Nutzers nur erhoben und verarbeitet werden, soweit dies erforderlich ist, um die Inanspruchnahme des Dienstes zu ermöglichen. Technisch erforderlich für die Interaktion und demnach zulässig ist die Verarbeitung einiger dieser Daten wie z. B. der IP-Adresse des Nutzerrechners oder der Ziel- und Startseiten beim Navigieren innerhalb des Angebotes. Der Einsatz eines vom Betreiber gesetzten Cookies kann je nach Funktionalität des Systems erforderlich sein, etwa wenn sich das Beibehalten von Einstellungen nicht anders realisieren lässt. Die Bildung eines gegebenenfalls personenbezogenen Profils oder der sonstige Einsatz von Data-Mining-Verfahren auf Nutzungsdaten zur Realisierung von individuellen Empfehlungen oder auf den einzelnen Nutzer zugeschnittenen Rankings ist nur dann zulässig, wenn diese Funktionen zu dem mit dem Nutzer vereinbarten Nutzungsinhalt gehören. Wann eine Funktion als vereinbart anzusehen ist, kann im Einzelfall schwierig zu bestimmen sein. Solange es sich jedoch nicht um eine für den nicht vorgebildeten Durchschnittsnutzer typische Grundfunktion handelt, muss ihm zumindest bei der Registrierung erläutert werden, welche Funktionen das System bietet. Denn nur dann kann der Nutzer einschätzen, ob er sich darauf einlassen möchte, mögen die genannten Datenverarbeitungsschritte auch ausschließlich im mutmaßlichen Interesse der Nutzerfreundlichkeit bestehen. Andere, nicht automatisch bei der Nutzung anfallende oder nicht für die unmittelbare Interaktion mit dem System erforderliche Daten, wie sie sich etwa aus einer Klickdatenanalyse ergeben, sind hingegen, sofern sie nicht ebenfalls der Durchführung einer vereinbarten Funktion dienen, von der Vorschrift nicht erfasst und dürfen daher ohne ausdrückliche Einwilligung des Nutzers nicht verwendet werden. Etwas anderes gilt nur, wenn technisch sichergestellt werden kann, dass eine Zuordnung zur Person oder zum Konto des Nutzers ausgeschlossen ist.

Aus den genannten und in ihrer Verwendung für die Interaktion zulässigen Nutzungsdaten darf der Betreiber nach § 15 Abs. 3 TMG Nutzungsprofile für Zwecke der Werbung oder zur bedarfsgerechten Gestaltung des Dienstes bilden, sofern er diese unter einem Pseudonym zusammenfasst und technisch sicherstellt, dass die Profildaten mit den Identifizierungsdaten nicht zusammengeführt werden. Der Diensteanbieter muss den Nutzer allerdings vorab über die Profilbildung informieren und ihn auf sein Widerspruchsrecht hinweisen. Widerspricht der betroffene Nutzer, darf der Anbieter keine Profilbildung vornehmen.

*Handlungsanweisung:* Der Betreiber darf personenbezogene Nutzungsdaten nur erheben und verwenden, wenn und soweit dies für die Durchführung des Dienstes erforderlich ist, was wiederum von dem vereinbarten Inhalt des Nutzungsverhältnisses abhängt. Über den grundsätzlichen Betrieb hinausgehende Funktionen müssen dem Nutzer daher bei der Registrierung erklärt werden. Aus diesen Daten dürfen auch zu Werbezwecken oder zur bedarfsgerechten Gestaltung Nutzungsprofile gebildet werden, jedoch nur unter Verwendung eines Pseudonyms und sofern der Nutzer dem nicht widerspricht. Der Nutzer ist über sein Widerspruchsrecht zu informieren. Sonstige Daten wie z. B. Klickdaten zur Erstellung von Rankings dürfen verwendet werden, wenn der Nutzer darin einwilligt oder der Betreiber technisch sicherstellt, dass die Verwendung anonym stattfindet, er also keinen Bezug zur Person des Nutzers herstellen kann.

### **Weiterreichung von Daten an Dritte**

BibSonomy bietet eine Anwendungs-Programmierschnittstelle (API) an, welche die Integration mit anderen Systemen ermöglicht, die BibSonomy-Daten über die API anfragen und verändern können. Dazu nutzt das Konzept der REST-API die typischen HTTP-Verben GET, PUT, POST und DELETE, um verschiedene Aktionen auf den URLs auszuführen. So ist es möglich, über die Anfrage `/api/tags` die Liste aller Schlagwörter des Systems zu erfragen. Die Anfrage <http://www.bibsonomy.org/api/posts?resourcetype=bookmark&search=folksonomy> liefert beispielsweise Einträge, bei denen der Begriff *folksonomy* im Titel, in der Beschreibung oder in den Tags vorkommt. Die in XML ausgegebenen Informationen enthalten unter Umständen auch personenbezogene Daten, wie

beispielsweise einen eindeutig auf eine bestimm- bare Person zurückzuführenden Benutzernamen. BibSonomy-Interessierte können solche Daten al- lerdings auch über die Weboberfläche in beliebigen Formaten (auch in XML) betrachten und für ihre Zwecke exportieren.

Der Abruf von zum Abruf bereitgehaltenen Daten durch (registrierte) Dritte, wie im Beispiel der API, stellt eine Datenübermittlung dar. Die Übermittlung von personenbezogenen Daten auf diesem Wege ist zulässig, wenn sie zu dem mit dem Nutzer vereinbarten Funktionsumfang gehört, sei es durch explizite Vereinbarung oder aufgrund der Tatsache, dass es sich dabei um eine typische Funktion des spezifischen Systems handelt, die den angesprochenen Durchschnittsnutzern selbstver- ständlich bekannt ist. In Systemen, in denen diese Funktion für den Nutzer unerwartet ist, ist eine Einwilligung des Nutzers in die Übermittlung erfor- derlich. Offene Programmierschnittstellen können als zentraler Bestandteil von Web2.0-Systemen, zu denen auch soziale Tagging-Systeme gehören, angesehen werden. Wenn solche Anwendungen zusätzlich noch in ihrem Funktionsumfang die Weitergabe ihrer öffentlichen Inhalte anbieten (zum Beispiel im Fall BibSonomy Publikationsdaten über verschiedene Exportfunktionen für die Erstellung von Referenzlisten), ist die Übermittlung mithilfe einer API zulässig.

*Handlungsanweisung:* Wenn die API zum ver- einbarten Funktionsumfang des Systems gezählt werden kann, ist die Bereitstellung einer solchen unproblematisch. Ansonsten können Betreiber die gesammelten Daten nur mit Einwilligung der jeweiligen Nutzer übermitteln oder müssen Tech- niken einsetzen, die die zu übermittelnden Daten endgültig und erfolgreich so anonymisieren, dass eine Zuordnung zu einem Nutzerkonto nicht mehr möglich ist.

### **Beendigung der Mitgliedschaft**

Möchte ein Benutzer nicht mehr Mitglied des Systems sein, so kann er sein Konto löschen. Auf Löschbefehl des Nutzers hin sperrt BibSonomy den entsprechenden Nutzerzugang, sodass die Einträge für die Öffentlichkeit und den Nutzer selber nicht mehr sichtbar sind und der Nutzer sich nicht mehr einloggen kann.

Mit der Beendigung des Nutzungsverhältnisses entfällt der Hauptzweck für die Speicherung und

anderweitige Verarbeitung sämtlicher personen- bezogener Daten des Nutzers. Bestandsdaten sind dementsprechend zu löschen. Unabhängig von der im Einzelfall bereits aus urheberrechtlichen Grün- den bestehenden Löschungspflicht von bestimmten Inhalten wie Fotos oder Videos bedarf es für die wei- tere Verwendung von Inhaltsdaten der Einwilligung des Nutzers oder nach den einschlägigen Vorschrif- ten des § 28 BDSG eines berechtigten Interesses des Anbieters. Weitere Voraussetzung ist, dass kein entgegenstehendes überwiegendes Interesse des Nutzers besteht. Die Löschung des Kontos ist jedoch ein Indiz dafür, dass der Nutzer seine „persönliche“ Beziehung zum Anbieter beenden und keine auf ihn zurückführbare Datenmasse zur freien Disposition des Anbieters hinterlassen will. Sofern technisch und organisatorisch nicht sichergestellt werden kann, dass der Bezug zur Person des Nutzers endgültig ausgeschlossen ist, sind die Inhaltsdaten demnach zu löschen. Etwaige Nutzungsdatenprofile unter Pseudonym sind ebenfalls zu löschen oder endgültig zu anonymisieren, da der Löschbefehl des Nutzers insoweit als Ausübung seines Widerspruchsrechts zu werten ist. Der Nutzer wird im Zweifel nicht davon ausgehen, dass dafür eine weitere Erklärung erforderlich ist.

*Handlungsanweisung:* Nach Löschung des Nut- zerkontos dürfen keine personenbezogenen Daten mehr verbleiben: sie sind zu löschen oder, sofern möglich, zu anonymisieren. Das heißt, Identifika- tionsmerkmale wie Realname oder E-Mailadresse und angesichts ihrer umstrittenen Einordnung auch die IP-Adresse sind zu löschen und der Rest einzeln oder unter einem für den Anbieter nicht mehr auf den Nutzer zurückzuführendem Pseud- onym abzuspeichern und zu verwenden. Bietet ein solches Restprofil dennoch die Möglichkeit, auf die ursprünglich dahinter stehende Person zu schließen, müssen die Daten gelöscht werden.

### **BibSonomy als Forschungsprojekt**

BibSonomy ist ein von einer Forschungseinrich- tung betriebenes System, welches dazu dient, die gewonnenen Erkenntnisse in der Praxis zu erpro- ben. Die Besonderheit der Forschung rund um Tagging-Systeme liegt in der leichtgewichtigen Wissensrepräsentation durch eine Folksonomy: Jeder Nutzer beschreibt seine Ressourcen individuell mit frei gewählten Tags. Durch Überlappung der Ressourcen und/oder Tags entstehen Beziehungen

zwischen Tags, Benutzern und Ressourcen. Die Analyse dieser Beziehungen wird für eine Reihe von Funktionalitäten benötigt. Dazu gehören die Spamentdeckung und das Ranking (siehe obere Abschnitte). Weitere Forschungsthemen sind die Strukturierung des Taggingvokabulars (z. B. durch eine automatische Extraktion von Synonym-/Hyponym-Beziehungen), Empfehlungssysteme (z. B. Tag-Empfehlungen für einen neuen Eintrag anhand der Tags von Nutzern, die ähnliche Tags und/oder Ressourcen nutzen) oder Clustering-Verfahren für die Entdeckung von Nutzergruppen.

Der Entwurf von Algorithmen für die genannten Aufgaben erfolgt Hand in Hand mit der Weiterentwicklung des Systems. Oft ist dabei im Vorhinein noch nicht klar, welche Daten für die einzelnen Anwendungsbereiche von Nutzen sein können. So hat sich beispielsweise erst nach einigen Beobachtungen ergeben, dass sich Spam-E-Mailadressen von anderen Adressen unterscheiden. Es ist anzunehmen, dass auch kommerzielle Anbieter mit verschiedenen Daten und Algorithmen experimentieren.

Aufgrund der vielfachen Nachfrage anderer Forschungsinstitutionen wurde aus den gesammelten öffentlichen Einträgen, also nur Inhaltsdaten, ein Benchmark-Datensatz erstellt. Mit diesem können auch externe Wissenschaftler BibSonomy-Daten für ihre Forschungszwecke nutzen. Die Nutzerdaten wurden pseudonymisiert, indem jedem Nutzernamen eine ID zugeordnet wurde. So ist ein direkter Rückschluss von Einträgen in dem Datensatz auf Einträge eines bestimmten Benutzers im System für andere als den Anbieter nicht möglich. In einigen Konstellationen kann durch eine Zuordnung der Beiträge aus dem Datensatz auf die öffentlichen Einträge der dahinter stehende Nutzer identifiziert werden, etwa wenn ein einzelnes Nutzerkonto sehr spezifische Tags oder Einträge aufweist. Ohne den Einsatz von Techniken, die die Daten dahingehend ändern, dass der Bezug zu einem Nutzerkonto vom Empfänger nicht mit verhältnismäßigem Aufwand hergestellt werden kann, wären daher im Zweifel alle zum Abruf bereitgehaltenen Daten als personenbezogen zu behandeln. Interessenten an dem beschriebenen Datensatz müssen eine Lizenz unterschreiben, mit der sie u. a. versichern, dass der Datensatz nur für Forschungszwecke genutzt und nicht weitergereicht wird.

Der datenschutzrechtlich zulässige Betrieb von BibSonomy als Forschungsprojekt unterliegt den Vorgaben des hessischen Datenschutzgesetzes (HDSG), insbesondere § 33 HDSG. Dieses Gesetz verdrängt im Fall von BibSonomy bezüglich Bestands- und Nutzungsdaten die Datenschutzvorschriften des weniger spezifischen TMG, da der Betrieb des Telemediums selbst den Forschungsgegenstand darstellt und alle dort gewonnenen Daten zu wissenschaftlichen Zwecken anhand ihrer Wirkung im System untersucht werden. So dürfen Bestands- und Nutzungsdaten etwa über dasjenige Maß hinaus, das für nichtwissenschaftliche Anbieter maßgeblich ist, erhoben und erforscht werden, um Data-Mining-Verfahren bspw. für Recommender-Systeme, Spamentdeckungsverfahren oder Ranking-Algorithmen zu entwickeln. Die vom Nutzer veröffentlichten Einträge (Inhaltsdaten) unterliegen nach § 3 Abs. 4 HDSG keinen datenschutzrechtlichen Beschränkungen und dürfen uneingeschränkt erhoben und verwendet werden, solange sie vom Nutzer veröffentlicht oder zur Veröffentlichung bestimmt sind. Grundsätzlich sind Daten auch im Bereich der wissenschaftlichen Forschung zu anonymisieren oder zu pseudonymisieren, wo dies möglich ist. Daneben gilt auch im wissenschaftlichen Bereich das Verbot, Daten ohne bestimmten Forschungszweck auf Vorrat zu sammeln. Der einzelne (Forschungs-)Zweck muss bei der Erhebung feststehen, wobei es in der Natur der Forschung liegt, dass sich einzelne Fragestellungen verändern oder ausweiten können.

### **Fazit**

Personenbezogene Daten werden durch das Grundrecht auf informationelle Selbstbestimmung geschützt. Konkrete Vorschriften liefern für Inhaltsdaten das BDSG und für Nutzungs- und Bestandsdaten das TMG. Gerade bei neuartigen Internetentwicklungen, in unserem Fall der (freiwilligen) Veröffentlichung von teilweise personenbezogenen Daten in sozialen Tagging-Systemen, ist nicht immer eindeutig, ob und wie die Daten und deren Besitzer geschützt werden können. Dieser Artikel zeigte anhand einer interdisziplinären Untersuchung des Systems BibSonomy, dass ein solcher Schutz unter Berücksichtigung rechtlicher Rahmenbedingungen auch in sozialen Tagging-Systemen möglich ist.

Zusammenfassend gilt, dass eine Verarbeitung von Inhaltsdaten unproblematisch ist, wenn sie für

das Systemangebot (Vertragsverhältnis) erforderlich ist. Da Tagging-Systeme auf der sozialen Interaktion vieler Nutzer basieren, ist die Veröffentlichung dieser Einträge Hauptzweck des Systems. Fraglich ist, inwieweit Funktionalitäten wie die Spamentdeckung oder eine verbesserte Suchfunktion zu den dem Nutzer bekannten und von ihm gewollten Bestandteilen zählen. Ähnlich vorsichtig ist der Umgang mit Bestands- und Nutzungsdaten zu bewerten. Bestandsdaten dienen der Begründung und inhaltlichen Ausgestaltung eines Nutzungsverhältnisses, Nutzungsdaten einer Ermöglichung der Inanspruchnahme des Telemediums. Die Erhebung, die Speicherung sowie die weitere Verarbeitung und Nutzung solcher Daten sind also grundsätzlich nur für diese Zwecke zulässig, und auch nur, wenn sie für deren Umsetzung erforderlich sind. Dieses muss für jedes einzelne Datum geprüft werden.

Neben einer Unterlassung der Datenerhebung können verschiedene Maßnahmen getroffen werden, die Entwickler, Betreiber und Benutzer schützen. Für Klarheit kann ein Anbieter sorgen, indem er seine Nutzer mit eindeutigen Informationen zu den angebotenen Funktionen sowie Art und Zweck der Datennutzung versorgt und ggf. eine Einwilligung abfragt, wo die Verwendung der Daten nicht von den vereinbarten Funktionen erfordert wird. So könnte man Nutzer beispielsweise wählen lassen, für welche Funktionen welche Daten zur Verfügung gestellt werden. Außerdem sollten, soweit möglich, anfallende Daten anonymisiert und, wenn nicht weiter benötigt, gelöscht werden.

Die heutigen rechtlichen und technischen Maßnahmen richten sich allerdings nicht an Daten, die Nutzer freiwillig und gezielt in Web2.0-Anwendungen veröffentlichen. Die damit verbundenen Folgen für die „entblößten“ Nutzer

lassen sich heute schwer absehen. Zur Wahrung des Grundrechts auf informationelle Selbstbestimmung im Internet werden langfristig Gestaltungsaspekte eine Rolle spielen, die Nutzer vor einer unbedachten Datenpreisgabe schützen oder die Möglichkeit geben, öffentliche Daten zu entfernen. Verschiedene Maßnahmen werden in Theorie und Praxis schon diskutiert. So ermöglichen unterschiedliche Privatheitsstufen eine Beschränkung der Sichtbarkeit von Einträgen auf nur einen Teil der Öffentlichkeit. Eine Art automatische Löschfunktion, die nach einigen Jahren die Einträge (auf Nachfrage) löscht, könnte helfen, vergessene Daten zu entfernen oder unsichtbar zu machen. Ein Anreiz zur Implementierung solcher Funktionalitäten in der Zukunft hängt allerdings von der Nachfrage der Nutzer und den rechtlichen Vorgaben ab, die auf das „soziale“ Internet abgestimmt werden müssen.

## Literatur

1. Benz D, Hotho A, Jäschke R, Krause B, Mitzlaff F, Schmitz C, Stumme G (2010) The social bookmark and publication management system BibSonomy – a platform for evaluating and demonstrating Web 2.0 research. VLDB J 19(5)
2. BVerfGE 1. Senat (1983) BVerfGE 65, 1 Volkszählung Urteil des Ersten Senats vom 15. Dezember 1983
3. Cattuto C, Schmitz C, Baldassarri A, Servidio VDP, Loreto V, Hotho A, Grahl M, Stumme G (2007) Network properties of folksonomies. AI Commun J 20:245–262
4. Hotho A, Jäschke R, Schmitz C, Stumme G (2006) BibSonomy: a social bookmark and publication sharing system. In: Moor AD, Polovina S, Delugach H (eds) Proceedings of the First Conceptual Structures Tool Interoperability Workshop at the 14th International Conference on Conceptual Structures. Aalborg Universitetsforlag, Aalborg, pp 87–102
5. Lerch H, Krause B, Hotho A, Roßnagel A, Stumme G (2010) Social Bookmarking-Systeme – die unerkannten Datensammler – ungewollte personenbezogene Datenverarbeitung? Multimedia und Recht 7:454–459
6. Morrison PJ (2008) Tagging and searching: search retrieval effectiveness of folksonomies on the World Wide Web. Inf Process Manag 44:1562–1579
7. Schild H-H, Ronellenfitsch M, Arlt U, Dembowski B, Wellbrock R (Hrsg) (2009) Hessisches Datenschutzgesetz (HDSG). Kommunal- und Schul-Verlag, Wiesbaden
8. Simitis S (2006) Kommentar zum Bundesdatenschutzgesetz. Nomos, Frankfurt
9. Roßnagel A, Scholz P (2000) Datenschutz durch Anonymität und Pseudonymität – Rechtsfolgen der Verwendung anonymer und pseudonymer Daten. Multimedia und Recht 3:721–729