Datenschutzrechtliche Bewertung des Projekts „Open-Access-Statistik“

I. Sachverhalt

Die Niedersächsische Staats- und Universitätsbibliothek Göttingen, die Humboldt-Universität zu Berlin, die Universitätsbibliothek der Universität Stuttgart, die Saarländische Universitäts- und Landesbibliothek sowie die Verbundzentrale des Gemeinsamen Bibliotheksverbunds (GBV) der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und der Stiftung Preußischer Kulturbesitz (VZG) führen gemeinsam das Projekt „Open-Access-Statistik 2“ durch, dessen Ziel es ist, Informationen über die Zugriffe auf Open-Access-Dokumente zu erfassen, auszuwerten und daraus Rückschlüsse auf die Nutzung von Open-Access-Dokumenten zu ziehen. Dies soll zum einen anhand einer Auswertung der Nutzhäufigkeit von Dokumenten geschehen, die die Projektpartner z. T. selbst im Rahmen so genannter Repositorien auf eigenen Dokumentenservern zur Verfügung stellen, zum anderen durch Auswertung von Suchanfragen, die über so genannte Linkresolverserver durchgeführt werden, die die Einrichtungen betreiben. Ein Linkresolverdienst sucht anhand der vom Nutzer vorgegebenen Kriterien nach Publikationen, erzeugt eine Ergebnisliste, die ggf. weitere Metainformationen zu den gefunden Dokumenten enthält, und generiert schließlich eine Verknüpfung auf das Dokument. Ein Linkresolverdienst hat damit im Wesentlichen Suchmaschinenfunktionalität.

Um eine Statistik über die Nutzungshäufigkeit der einzelnen Dokumente erstellen zu können, soll jeder Dokumentenabruf bzw. jede Suchanfrage, die an die Dokumenten- und Linkresolverserver gerichtet werden, gespeichert werden. In den meisten Fällen protokollieren die Betreiber der Dienste die Nutzeraktivitäten in Logfiles, die der Webserver selbst erzeugt. Im Rahmen des Projekts „OA-Statistik 2“ sollen dabei die folgenden Daten gespeichert und ausgewertet werden: IP-Adresse des Nutzers, Tag und Uhrzeit der Ausführung, Dateipfad und Dateiname des angeforderten Dokuments, HTTP-Methode, HTTP-Statuscode, Größe des angeforderten Dokuments in Byte, übertragene Byte, Spezifikation des vom Nutzer eingesetzten Clients (User-Agent-Parameter im HTTP-Header der Nutzeranfrage), bibliotheksinterne Dokumenten-ID, Referrer-Angabe im HTTP-Header, ggf. Accept-Header.

Die Nutzungshäufigkeit soll dabei jedoch in mehrfacher Hinsicht um die Statistik verfälschende Faktoren bereinigt werden: die mehrmalige Nutzung eines Dokuments innerhalb einer bestimmten Zeitspanne (sog. COUNTER-Doppelclickspanne)[1] von einem Rechner aus soll als ein einziger Nutzungsvorgang interpretiert werden und als einfacher Download in die Zählung eingehen. Da identische Dokumente, die von mehreren Autoren verfasst wurden, auf unterschiedlichen Servern parallel veröffentlicht worden sein können, soll bei der Open-Access-Statistik der mehrmalige Abruf eines Dokuments in Form identischer, aber auf verschiedenen Servern liegender Dateien innerhalb der COUNTER-Doppelclickspanne von einem Arbeitsplatz aus als solcher identifiziert und ebenfalls als nur einmaliger Download gezählt werden. Außerdem soll der Abruf von an verschiedenen Einrichtungen vorgehaltenen Dokumenten von einem Arbeitsplatz aus als zusammengehörig identifiziert werden, um Downloadgraphen zu ermitteln und Empfehlungsdienste anbieten zu können, die dem Nutzer (ähnlich wie bei Amazon oder Ebay) Dokumentenempfehlungen anzeigen, die sich aus dem aktuell aufgerufenen bzw. gesuchten Dokument ergeben. Schließlich sollen Dokumentenaufrufe nicht menschlichen Ursprungs, also insbesondere durch Webcrawler, als solche identifiziert werden und bei der Berechnung der Nutzungshäufigkeit unberücksichtigt bleiben. Bei Webcrawlern handelt es sich um (häufig von Suchmaschinenbetreibern eingesetzte) Computerprogramme, die Internetseiten automatisiert durchsuchen, aus ihren Inhalten Indizes erzeugen und die aufgerufenen Webseiten in der Regel zumindest auszugsweise speichern. Hierfür rufen sie selbsttätig alle erreichbaren Internetseiten auf, und zwar auch frei zugängliche Dokumente digitaler Bibliotheken. Da die Open-Access-Statistik ausschließlich Dokumentenaufrufe natürlicher Personen erfassen soll, sollen solche automatisierten Dokumentenaufrufe durch Webcrawler erkannt und ausgefiltert werden.

Diese Daten sollen an einen zentralen Serviceprovider außerhalb der jeweiligen Einrichtung weitergegeben werden. Als Serviceprovider soll die VZG fungieren. Sie soll die entsprechenden Auswertungen vornehmen, also z. B. Mehrfachzugriffe innerhalb der COUNTER-Doppelklickspanne von gleichen Rechnern aus einrichtungsübergreifend durch Abgleich der Datensätze erkennen, und auf diese Weise eine Open-Access-Statistik erstellen, deren Informationen dann an die anderen Projektpartner zurückgeleitet werden, die die Informationen bei zukünftigen Dokumentenaufrufen und Suchanfragen anzeigen.

[1] Counter bezeichnet einen Zählstandard im Bereich E-Publikation, http://www.projectcounter.org/

II. Datenschutzrechtliche Bewertung

1. Anwendbares Recht

a) Personenbezogene Daten

Datenschutzrechtliche Bestimmungen sind nur anzuwenden, soweit **personenbezogene oder –beziehbare Daten** verarbeitet werden (§ 2 Abs. 1 S. 1 LDSG/NDSG[2]).
Die den Nutzern beim Abruf der Inhalte von den Repositorien zugewiesenen dynamischen und statischen **IP-Adressen** sind im vorliegenden Projekt als zumindest personenbeziehbar anzusehen ungeachtet der aktuellen Diskussion, ob IP-Adressen für Inhalteanbieter (Content-Provider) personenbeziehbar sind oder nicht.[3] Denn im vorliegenden Fall fungieren die Dataprovider häufig nicht nur als Inhalteanbieter, indem sie in den eigenen Repositorien Open-Access-Dokumente zum Abruf bereitstellen, sondern auch als Zugangsanbieter (Access-Provider), indem sie den eigenen Nutzern einen Internetzugang zur Verfügung stellen, wie es z. B. Hochschulen im Allgemeinen für ihre Mitglieder und Angehörigen tun. Es ist unstrittig, dass IP-Adressen jedenfalls für den Zugangsanbieter personenbezogene Daten sind, weil dieser selbst die IP-Adressen den Nutzern zuweist und die Zuordnung daher kennt. Es führen auch nicht sämtliche Dataprovider im Rahmen ihres Internetzugangsangebots eine Adressumsetzung (NAT) durch mit der Folge, dass alle Nutzer dieser Einrichtung mit derselben IP-Adresse auf ein Repositorium zugreifen. Wie die Gespräche im Projekt ergeben haben, werden Nutzern durchaus auch individuelle IP-Adressen zugewiesen.
Im Ergebnis sind die den Nutzern zugewiesenen IP-Adressen für die Dataprovider als personenbeziehbar anzusehen.
Im Zusammenhang mit der IP-Adresse sind alle weiteren von den Dataprovidern erfassten Daten ebenfalls als personenbezogen anzusehen.

b) Allgemeine Vorschriften

Die Frage, welches Recht anzuwenden ist, hängt davon ab, welche Stelle die personenbezogenen Daten (für sich selbst) verarbeitet oder durch andere im Auftrag verar-

[2] Der Einfachheit halber werden nicht immer alle Landesvorschriften aufgezählt. Da diese inhaltlich meist übereinstimmen, werden nur einige Vorschriften exemplarisch genannt.
[3] Die wohl überwiegende Ansicht bejaht einen Personenbezug, insbesondere die Aufsichtsbehörden: Beschluss der obersten Aufsichtsbehörden für den Datenschutz im nicht-öffentlichen Bereich am 26./27. November 2009, „Datenschutzkonforme Ausgestaltung von Analyseverfahren zur Reichweitenmessung bei Internet-Angeboten“ [abrufbar unter: www.lfd.m-v.de/dschutz/beschlue/Analyse.pdf]; ebenso Artikel-29-Datenschutzgruppe, Stellungnahme 4/2007 zum Begriff „personenbezogene Daten“, 20.06.07, S. 19 f.; zum Streitstand: Sachs, Datenschutzrechtliche Bestimmbarkeit von IP-Adressen, CR 2010, 547 ff.

beiten lässt (so genannte „verantwortliche Stelle“[4] oder „Daten verarbeitende Stelle“[5]) bzw. wer Diensteanbieter im Sinne des Telemediengesetzes ist.
Die Verarbeitung personenbezogener Daten richtet sich grundsätzlich nach den allgemeinen Datenschutzgesetzen (das Bundesdatenschutzgesetz sowie die Datenschutzgesetze der Länder), soweit nicht Spezialregelungen bestehen. Solche bereichsspezifischen Vorschriften verdrängen die allgemeinen Vorschriften (§ 2 Abs. 5 LDSG, § 2 Abs. 6 NDSG).
Für inländische nicht-öffentliche Stellen (z. B. Vereine, Gesellschaften mit begrenzter Haftung) und öffentliche Stellen des Bundes gilt i. d. R. das **Bundesdatenschutzgesetz** (§ 1 Abs. 2 BDSG), für öffentliche Stellen der Länder (insb. Körperschaften und Anstalten des öffentlichen Rechts wie Hochschulen, Landesbibliotheken u. ä.) das jeweilige **landesspezifische Datenschutzgesetz**, in dem die Stelle ihren Sitz hat (§ 1 Abs. 2 Nr. 2 BDSG i. V. m. Landesrecht). Für Stellen mit Niederlassung in anderen EU-Staaten gelten die datenschutzrechtlichen Bestimmungen des jeweiligen Staates.[6]

c) Bereichsspezifische Vorschriften

Bereichsspezifische Rechtsvorschriften, die diesen allgemeinen Bestimmungen vorgehen, enthält u. a. das Telemediengesetz (TMG)[7]. Dieses findet gemäß § 1 TMG auf alle **Telemedien** Anwendung, d. h. auf alle elektronischen Informations- und Kommunikationsdienste, soweit sie nicht Telekommunikationsdienste (bei denen ausschließlich Signale über Telekommunikationsnetze übertragen werden, ohne dass eine inhaltliche Dienstleistung erbracht wird), telekommunikationsgestützte Dienste oder Rundfunk sind. Es ist unerheblich, ob ein Entgelt für den Dienst erhoben wird, oder ob es sich bei dem Anbieter um eine öffentliche oder nicht-öffentliche Stelle handelt (§ 1 Abs. 1 S. 2 TMG).
Ein Dokumentenserver stellt ein derartiges inhaltliches Angebot im Sinne des § 1 TMG dar, weil elektronische Dokumente zum Abruf bereit gestellt werden; ein Telekommunikationsdienst, ein telekommunikationsgestützter Dienst oder Rundfunk liegt darin nicht. Dasselbe gilt für Linkresolver, also einen Dienst, der Suchabfragen speziell im Bibliotheksbereich durchführt, indem bibliographische Angaben zu Medien mit weiteren Informationen (z. B. zur Verfügbarkeit an einem Standort) verknüpft und Verknüpfungen auf den Volltext des gefundenen Dokuments erzeugt werden. Insofern handelt es sich um einen Online-Dienst, der Instrumente zur Datensuche und zum Zugang zu Daten bereitstellt.[8]

[4] So z. B. § 3 Abs. 3 LDSG.
[5] So z. B. § 3 Abs. 3 NDSG.
[6] Simitis, BDSG, § 1 Rn. 198 ff.
[7] Gesetz vom 26. Februar 2007, BGBl. I S. 179, zuletzt geändert durch Art. 1 1. Telemedienänderungsgesetz vom 31. 5. 2010.
[8] Siehe BT-DRs. 16/3078, S. 13.

Verantwortlich für die Einhaltung der Vorgaben des TMG ist der so genannte „**Diensteanbeiter**“. Darunter ist nach § 2 Nr. 1 TMG jede natürliche oder juristische Person zu verstehen, die eigene Telemedien zur Nutzung bereithält oder den Zugang zur Nutzung vermittelt. Indem die Bibliotheken einen Webserver bzw. Dokumentenserver betreiben, halten sie eigene bzw. fremde Telemedien zur Nutzung bereit. Bei Linkresolvern liegt neben einer eigenen inhaltlichen Angebotsleistung ggf. eine Zugangsvermittlung vor. Damit sind die Universitäten und Landesbibliotheken grundsätzlich Diensteanbieter i. S. d. § 2 Nr. 1 TMG. Auf sie findet hinsichtlich der Datenverarbeitung bereichsspezifisch das TMG Anwendung, soweit sie ihren Sitz in Deutschland haben (§ 3 Abs. 3 Nr. 4 TMG i. V. m. § 1 BDSG).[9]

Die spezialgesetzlichen *Vorschriften über den Datenschutz* im TMG (§§ 11 ff. TMG) gelten jedoch nur im **Anbieter-Nutzer-Verhältnis**. An einem solchen fehlt es, wenn der Dienst im Dienst- und Arbeitsverhältnis zu ausschließlich beruflichen oder dienstlichen Zwecken bereitgestellt wird (§ 11 Abs. 1 Nr. 1 TMG). Entsprechendes wird anzunehmen sein, wenn der Dienst im Rahmen des Studiums ausschließlich zu studienbezogenen Zwecken genutzt werden darf.[10] Dies trifft auf die hier relevanten Telemediendienste nicht zu. Bereits der Open-Access-Ansatz bedingt, dass die gespeicherten Dokumente frei zugänglich sind, so z. B. auch über Suchmaschinen auffindbar sind und durch jedermann abgerufen werden können, das Dienstangebot daher auch zu privaten Zwecken genutzt werden kann. Auch die Linkresolverdienste, die u. U. auf Closed-Access-Angebote verweisen, sind selbst frei nutzbar. Selbst eine etwaige gemischte Nutzung (sowohl privat als auch dienstlich/studienbezogen) führt dem Wortlaut nach („ausschließlich“) für den gesamten Dienst zwingend zur Anwendung der §§ 11 ff. TMG.[11] Etwas anderes wäre lediglich dann denkbar, wenn sich das Telemedienangebot getrennt betrachten lässt (Aufspaltung in private und dienstliche/studienbezogene Nutzung) zum Beispiel indem getrennte Nutzerkonten für die private und dienstliche/studienbezogene Nutzung existieren oder Anfragen über unterschiedliche Eingabemasken abgewickelt werden. Dies ist vorliegend jedoch nicht der Fall. Damit sind die datenschutzrechtlichen Vorschriften des TMG einschlägig.

[9] Pfeiffer/Weller/Nordmeier, in: Spindler/Schuster, Recht der elektronischen Medien, § 3 Rn. 1; Spindler/Nik, aaO, § 11 TMG Rn. 15.
[10] Diese Auffassung wurde bereits zur gleichlautenden Vorgängerregelung im Teledienstedatenschutzgesetz (TDDSG) vertreten: Der Hessische Landesbeauftragten für den Datenschutz, 28. Tätigkeitsbericht, 1999, Kapitel 9.1.1.1; Gola/Müthlein, TDG/TDDSG, 2000, § 2 TDDSG, Rn. 3.4. und 3.5.
[11] Spindler/Nink, in: Spindler/Schuster, Recht der elektronischen Medien, TMG § 11 Rn. 11.

2. Verantwortlichkeit

Verantwortlich für die Einhaltung der datenschutzrechtlichen Bestimmungen ist der Diensteanbieter bzw. die verantwortliche Stelle. Das ist nicht die mit der Datenverarbeitung befasste Funktionseinheit (wie etwa eine Hochschulbibliothek oder ein Hochschulrechenzentrum), sondern die betreffende rechtlich selbständige juristische Person (also z. B. die jeweilige Hochschule als Körperschaft des öffentlichen Rechts oder eine Landesbibliothek als rechtsfähige Anstalt).[12] Dementsprechend sind an dem vorliegenden Projekt verschiedene verantwortliche Stellen beteiligt:

(a) Die **Dataprovider**: Der Begriff Dataprovider ist hier nicht im Sinne eines Computerprogramms oder einer Anwendung zu verstehen, sondern als Bezeichnung derjenigen Diensteanbieter, die einen Dokumenten- oder Linkresolverserver betreiben und dabei die im Rahmen des OAS-Projekts verarbeiteten Daten erheben und an den zentralen Serviceprovider weiterleiten. Dataprovider in diesem Sinn sind z. B. die Georg-August-Universität Göttingen, die Universität Stuttgart, die Universität des Saarlands und die Humboldt-Universität zu Berlin. Mit dem Fortgang des Projekts ist damit zu rechnen, dass weitere in- und ggf. auch ausländische Dataprovider hinzukommen.

(b) Der **Serviceprovider**: mit Serviceprovider ist diejenige verantwortliche Stelle gemeint, an die die Dataprovider die erhobenen Daten zentral weiterleiten. Der Serviceprovider wertet diese Daten aus, bereitet sie auf und meldet das Ergebnis schließlich an die einzelnen Dataprovider zurück. Diese Aufgabe übernimmt die VZG, die Verbundzentrale des Gemeinsamen Bibliotheksverbunds (GBV) der Länder Bremen, Hamburg, Mecklenburg-Vorpommern, Niedersachsen, Sachsen-Anhalt, Schleswig-Holstein, Thüringen und der Stiftung Preußischer Kulturbesitz. Die VZG ist als Dienstleistungszentrum des GBV gem. § 4 Abs. 1 des Verwaltungsabkommens über die Errichtung eines Bibliotheksverbundes[13] hinsichtlich der fachlichen Aufsicht von der Staats- und Universitätsbibliothek Göttingen unabhängige Einrichtung des Landes Niedersachsen (vgl. insofern auch § 4 Abs. 2 des Verwaltungsabkommens). Sie ist ein niedersächsischer Landesbetrieb nach § 26 der Niedersächsischen Landeshaushaltsordnung (LHO).[14] Insofern ist sie eine im datenschutzrechtlichen Sinne eigenständige Daten verarbeitende Stelle i. S. d. § 3 Abs. 3 NDSG.[15]

[12] Dammann, in: Simitis, BDSG, § 3 Rn. 225; Gola/Schomerus, BDSG, § 3 Rn. 48; Artikel-29-Datenschutzgruppe, Stellungnahme 1/2010 zu den Begriffen „für die Verarbeitung Verantwortlicher“ und „Auftragsverarbeiter“, WP 169, 19 ff.
[13] abrufbar unter: http://www.gbv.de/bibliotheken/allgemeines/gemeinsamer-bibliotheksverbund-gbv/02GBV_1200
[14] http://www.gbv.de/bibliotheken/allgemeines/VZG/ueber_die_VZG/index
[15] Vgl. auch Simitis, BDSG, § 3 Rn. 231, wonach einzelne Landesbehörden eigenständige verantwortliche Stellen sind.

Weitere Akteure in dem Verfahren sind schließlich die **Nutzer**. Nutzer ist nach § 11 Abs. 2 TMG diejenige natürliche Person, die Telemedien nutzt, insbesondere um Informationen zu erlangen oder zugänglich zu machen. In der allgemeinen datenschutzrechtlichen Terminologie sind sie Betroffene (§ 3 Abs. 1 LDSG/NDSG).

3. Verarbeitung von Nutzungsdaten

a) Grundsatz: präventives Verbot mit Erlaubnisvorbehalt

Im Projekt OA-Statistik sollen die Dataprovider unterschiedliche Informationen, die beim Dokumentenabruf bzw. der Abwicklung von Rechneranfragen vom Webserver verarbeitet werden, zumindest temporär speichern und an den Serviceprovider weiterleiten.
Im Datenschutzrecht besteht ein präventives Verbot mit Erlaubnisvorbehalt: jegliche Verarbeitung personenbezogener Daten ist **grundsätzlich verboten, es sei denn sie ist durch eine Rechtsvorschrift oder durch die Einwilligung des Betroffenen gestattet**, § 4 Abs. 1 LDSG/NDSG. Dies gilt auch für Telemediendienste (§ 12 Abs. 1 TMG). Demnach dürfen die Dataprovider personenbezogene Daten beim Betrieb von Dokumenten- und Linkresolverservern grundsätzlich nur erheben und verwenden, soweit das TMG oder eine andere Rechtsvorschrift mit ausdrücklichem Bezug auf Telemedien dies erlaubt oder der Nutzer einwilligt. Auf die Einwilligung als Rechtsgrundlage soll im vorliegenden Projekt nicht zurückgegriffen werden. Sie wird daher im Folgenden nicht weiter erwähnt.

b) Datenverarbeitung für die technische Abwicklung von Client-Anfragen

Bei der Verarbeitung von Nutzungsdaten ist zu unterscheiden zwischen der Datenverarbeitung zum Zweck der reinen Diensterbringung (also der Anzeige von Dokumenten bzw. der Durchführung von Suchabfragen), und der (weiteren) Verarbeitung der Daten zu anderen Zwecken, wie sie im Projekt OA-Statstik zur Ermittlung von Zugriffszahlen erfolgt.

Nach § 15 Abs. 1 TMG dürfen die Dataprovider als Diensteanbeiter i. S. d. TMG personenbezogene Daten der Nutzer nur erheben und verwenden, **soweit es erforderlich ist, die Inanspruchnahme der Telemedien zu ermöglichen** und abzurechnen[16]. Diese Daten definiert das Gesetz als Nutzungsdaten. Nutzungsdaten sind nach § 15 Abs. 1 TMG insbesondere Merkmale zur Identifikation des Nutzers, Angaben über Beginn

[16] Eine Verarbeitung personenbezogener Daten zu Abrechnungszwecken dürfte angesichts der Bereitstellung von Open-Access-Inhalten nicht in Betracht kommen.

und Ende sowie des Umfangs der jeweiligen Nutzung und Angaben über die vom Nutzer in Anspruch genommenen Telemedien. Mit anderen Worten sind Nutzungsdaten sämtliche Informationen, die bei der Interaktion zwischen Nutzer und Anbieter während und durch die Dienstenutzung notwendigerweise verarbeitet werden.[17] Damit ist ein strenger Maßstab bei der Frage anzulegen, wann eine Datenverarbeitung erforderlich ist, um die Inanspruchnahme von Telemedien zu ermöglichen. Nicht erforderliche Daten dürfen nämlich schon nicht vom Webserver erhoben und folglich auch nicht zweckändernd genutzt werden.[18] Bereits die Erhebung von Daten, also das (auch nur temporäre) Erfassen von Daten durch den Webserver, muss sich auf das für die Erbringung des Dienstes unbedingt erforderliche Maß beschränken.[19] Vor diesem Hintergrund sind die von den Dataprovidern im Rahmen von OA-Statistik verarbeiteten Daten zu bewerten.

§ 15 Abs. 1 TMG rechtfertigt die Verarbeitung jedenfalls derjenigen Daten, die **technisch für den Dienstbetrieb zwingend notwendig** sind. Das sind zum einen diejenigen Informationen, die der Client, also der Webbrowser des Nutzers, zum Zweck der Kommunikationsabwicklung an den Webserver des Dataproviders sendet; außerdem die Daten, die der Webserver dienstnotwendig im Rahmen des Kommunikationsvorgangs selbst erzeugt.
Zunächst sendet der Client eine Anfrage an den Server. Diese enthält auf Transportebene die IP-Adresse des Clients, die der Webserver schon deswegen zwingend verarbeiten muss, um die Rechneranfrage überhaupt beantworten zu können. Die Anfrage des Clients enthält außerdem einen so genannten Accept-Header mit Informationen darüber, welche Inhaltstypen, Zeichensätze, Kodierungen und Sprachen vom Client bevorzugt bzw. ausschließlich verarbeitet werden. Auch diese Informationen sind damit für die Diensterbringung erforderlich, weil sie spezifizieren, wie der Webserver die Anfrage beantworten soll oder muss. Außerdem sendet der Client in der Anfrage einen User-Agent-Header an den Server, der Angaben zum eingesetzten Client-Programm enthält. Diese Informationen können für eine „Browserweiche“ erforderlich sein: da nicht alle Client-Programme gleichermaßen mit allen verfügbaren Skriptsprachen und Protokollen umgehen können, kann es für den Webserver erforderlich sein, den eingesetzten Useragent zu erkennen, um eine einheitliche bzw. richtige Darstellung der Inhalte sicherzustellen.
Des Weiteren müssen sich Client und Webserver auf ein gemeinsames Protokoll verständigen; dies macht den Austausch von Informationen über Protokollversion und Request-Methode, also auf welche Weise Daten übertragen werden, erforderlich. In seiner Antwort sendet der Webserver automatisch die Größe des zu übertragenden Ob-

[17] Spindler/Nink, in: Spindler/Schuster, Recht der elektronischen Medien, TMG § 15 Rn. 2.
[18] Spindler/Nink, in: Spindler/Schuster, Recht der elektronischen Medien, TMG § 11 Rn. 5.
[19] Schmitz, in: Spindler/Schmitz/Geis, TDG, § 6 TDDSG Rn. 8.

jekts an den Client (hier: Größe des Dokuments in Byte) und verarbeitet die Information wie viele Bytes übertragen wurden. Notwendigerweise muss der Webserver außerdem den Dateipfad (URL) und den Dateinamen des abgerufenen Dokuments verarbeiten, um die gewünschte Seite bzw. das gewünschte Dokument anzeigen zu können. Außerdem erzeugt und sendet der Webserver auf jede Client-Anfrage einen HTTP-Statuscode, der dem anfragenden Rechner notwendige Informationen zum Stand der Anfrage gibt (z. B. ob und welcher Server- oder Client-Fehler festgestellt wurde). Des Weiteren werden vom Webserver standardmäßig Datum und Uhrzeit einer Anfrage bzw. der Antwort des Servers verarbeitet. Diese Information fällt daher notwendigerweise bei der Dienstnutzung an.

Informationen über die verweisende Verknüpfung (den so genannten „Referrer“, also den Link, über den der Nutzer auf einen Inhalt des Webservers gelangt ist) sind dagegen nicht erforderlich für die technische Abwicklung von Client-Anfragen. Es handelt sich jedoch um vom Browser häufig, aber nicht in allen Fällen übermittelte Informationen zum Ursprungsort. Der Referrer wird vom Client automatisch ohne weiteres Zutun des Webservers übertragen und von diesem zwangsläufig „zur Kenntnis genommen“. Dieser Vorgang bedarf deshalb noch keiner datenschutzrechtlichen Rechtfertigung. Etwas anderes gilt für die weitere Verarbeitung durch den Webserver, also insbesondere die Speicherung des Referrers in Nutzungsprofilen (dazu unten).

Zwischenergebnis: Bis auf den Referrer sind sämtlich im Rahmen von OA-Statistik von den Webservern erfasste Daten für die Diensterbringung technisch zwingend notwendig und werden deshalb von den Webservern (temporär) zulässigerweise auf Grundlage des § 15 Abs. 1 TMG verarbeitet. Der Referrer selbst wird dem Webserver ohne Weiteres Zutun automatisch zugesandt.

c) Weitere Datenverarbeitung für OA-Statistik

Die genannten Daten werden zunächst vom Webserver verarbeitet und anschließend in aller Regel in einem Webserverlog gespeichert. Bereits diese Speicherung ist für den technischen Betrieb der Webserver nicht mehr zwingend notwendig, so dass sich die Frage stellt, ob dieser Verarbeitungsschritt noch auf § 15 Abs. 1 TMG (Ermöglichung der Inanspruchnahme von Telemedien) gestützt werden kann. Dieselbe Frage stellt sich für die sich anschließenden Verarbeitungsschritte. So werden im Projekt OA-Statistik die erfassten Nutzungsdaten von einer Anwendung aus dem Webserverlogfile ausgelesen und anschließend über OAI-PMH an den Webserver des Serviceproviders übertragen.

Welche Datenverarbeitung noch der „Ermöglichung der Inanspruchnahme von Telemedien“ dient, wird in der Literatur primär vom Sinn und Zweck des konkreten Dienstes abhängig gemacht, so dass entscheidend ist, ob der Sinn und der Zweck des Dienstes die Datenverarbeitung erfordert.[20] Dabei wird an die Erforderlichkeit jedoch ein sehr strenger Maßstab angelegt.[21] So soll die Speicherung und Anzeige von Grunddaten der Nutzer in sozialen Netzwerken auf § 15 Abs. 1 TMG gestützt werden können, weil es gerade Zweck eines sozialen Netzwerks sei, die Nutzer untereinander auffindbar zu machen. Dagegen lasse sich die Speicherung und Anzeige von Informationen wie Hobbys, Interessen oder den beruflichen Werdegang nicht auf § 15 Abs. 1 TMG stützen, weil dies für die Inanspruchnahme des Netzwerks nicht unerlässlich sei.

Überträgt man dies auf das Projekt OA-Statistik, so dürfte die Frage, ob die bei den Repositorien erfolgende Datenverarbeitung für OA-Statistik für die Inanspruchnahme der Telemedien erforderlich ist, zu verneinen sein. Sinn und Zweck eines Dokumentenservers ist es, dem Nutzer elektronische Dokumente zur Verfügung zu stellen, die Dokumente zu archivieren und über Metadaten erschließbar zu machen. Dafür ist es aber nicht erforderlich, das Nutzerverhalten auszuwerten und Zugriffsstatistiken zu erstellen. Dies gilt auch für Linkresolver. Die mit OA-Statistik bezweckten Auswertungen sollen den Nutzern vielmehr einen Mehrwert bzw. einen Zusatznutzen bieten, der zwar die Inanspruchnahme der Dienste erleichtert und ggf. optimiert, aber nicht erst ermöglicht.
Für eine solche Betrachtung sprechen auch systematische Erwägungen. So hat der Gesetzgeber in § 15 Abs. 3 TMG eine Rechtsgrundlage für Datenverarbeitungen geschaffen, die der „bedarfsgerechten Gestaltung der Telemedien“ dienen, also auch von Maßnahmen, die der Ausrichtung des Angebots am Bedarf der Nutzer dienen und damit der Optimierung des Telemediums. Dieser Erlaubnistatbestand wäre überflüssig, wenn sich die Zulässigkeit einer entsprechenden Datenverarbeitung bereits aus § 15 Abs. 1 TMG ergeben würde.

Folglich kann die Speicherung von Nutzungsdaten in Webserverlogfiles sowie deren weitere Verarbeitung beim Dataprovider nicht auf § 15 Abs. 1 TMG gestützt werden.[22]

---

[20] Spindler/Nink, in: Spindler/Schuster, Recht der elektronischen Medien, TMG § 15 Rn. 5a.
[21] a. A. Zscherpe, in: Taeger/Gabel, Kommentar zum BDSG, 2010, TMG § 15 Rn. 29, jedoch mit dem ausdrücklichen Hinweis auf die anderslautende herrschende Meinung (aaO, FN 35).
[22] So im Übrigen auch schon Amtsgericht Mitte (Berlin) vom 27.03.07 (Az. 5 C 314/06); noch strenger, aber im Ergebnis nicht überzeugend: Datenschutzkonforme Ausgestaltung von Analyseverfahren zur Reichweitenmessung bei Internet-Angeboten, Beschluss der obersten Aufsichtsbehörden für den Datenschutz im nicht-öffentlichen Bereich vom 26./27.11.2009; Landesbeauftragter für den Datenschutz Bayern, 23. Tätigkeitsbericht (2007/2008), LT-Drs. 16/2100, S. 23 f; Gestaltung des Internetauftritts, Landesbeauftragter für den Datenschutz Bayern, http://www.datenschutz-bayern.de/big/technik/orient/internetauftritt.html; Anonymisierung der IP-Adressen in Webserver-Logfiles, Sächsischer Datenschutzbeauftragter, http://www.saechsdsb.de/ipmask. Wie hier: FAQ IP-Adressen und andere Nutzungsdaten, Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein, https://www.datenschutzzentrum.de/ip-adressen/.

d) Rechtsgrundlagen für die weitere Verarbeitung

Nach § 12 Abs. 2 TMG dürfen die erhobenen Nutzungsdaten daher ohne Einwilligung der Nutzer nur dann in einem Webserverlogfile gespeichert und von den Dataprovidern für OA-Statistik weiterverarbeitet werden, soweit eine andere Rechtsvorschrift, die sich auf Telemedien bezieht (insbesondere im TMG), es erlaubt oder die Nutzer einwilligen. Andernfalls greift der Grundsatz, dass im Rahmen der Bereitstellung von Telemedien angefallenen personenbezogenen Daten unmittelbar nach Beendigung der Nutzung gelöscht werden müssen (§ 13 Abs. 4 Nr. 2 TMG).

aa) Zulässige Zwecke des § 15 Abs. 3 TMG

Eine Rechtsvorschrift, die eine weitere, zweckändernde Verwendung der Nutzungsdaten auch ohne Einwilligung der Betroffenen gestattet, ist § 15 Abs. 3 TMG. Danach dürfen aus den vom Webserver erhobenen Daten für die Zwecke **Werbung, Marktforschung** sowie **bedarfsgerechte Gestaltung der Telemedien** pseudonyme Nutzungsprofile erstellt werden. Eine gesetzliche Definition der Begriffe gibt es nicht, sie sind daher auslegungsbedürftig.[23]

Im Projekt OA-Statistik kommt als Rechtsgrundlage in erster Linie die Verarbeitung der Nutzungsdaten zum Zweck der bedarfsgerechten Gestaltung der Telemedien in Betracht. Darunter dürften alle Maßnahmen fallen, die aus Sicht der Nutzer der Optimierung des Diensteangebots dienen.
Im Projekt OA-Statistik sollen letztendlich Zugriffsstatistiken für einzelne Dokumente generiert und den Nutzern in Form von Trefferlisten, Dokumentenempfehlungen oder Dokumentenrankings angezeigt werden, die als nutzungsbasierte Bewertung Aufschluss über die Nutzungshäufigkeit und damit über die qualitative und quantitative Bedeutung eines Dokuments geben. Jedenfalls durch diese Funktion wird den Nutzern neben der bloßen Katalogisierung und Bereitstellung von Dokumenten eine weitere Dienstleistung angeboten, die insgesamt die Suche nach Dokumenten verbessern und damit das inhaltliche Angebot der Dokumentenserver und Linkresolver im Allgemeinen optimieren soll. Diese Maßnahme dient der bedarfsgerechten Gestaltung der Telemedien.

Fraglich ist, ob die darüber hinaus verfolgten Ziele wie etwa das, „die Akzeptanz von Open-Access-Dokumenten unter Autoren und Rezipienten zu erhöhen“[24] unter die in § 15 Abs. 3 TMG genannten Zwecke Werbung oder Marktforschung noch subsumiert

[23] Schmitz, in: Spindler/Schmitz/Geis, § 6 TDDSG Rn. 27.
[24] Projektantrag „Dienste und Standards für international vergleichbare Nutzungsstatistiken (Fortsetzungsantrag), Kennwort: OA-Statistik 2, S. 5.

werden können.[25] Letztendlich kann diese Frage jedoch dahinstehen, da vorliegend jedenfalls eine bedarfsgerechte Gestaltung der Telemedien erfolgen soll. Damit wird bei der Datenverarbeitung durch die Dataprovider zumindest ein erlaubter Zweck verfolgt.

In Bezug auf den Referrer wird die Auffassung vertreten, dass dessen Speicherung und Verwendung für pseudonymisierte Nutzungsprofile dann zulässig ist, wenn die verantwortliche Stelle zweifelsfrei nachweisen kann, dass die Erstellung der Profile für Werbung, Marktforschung oder die bedarfsgerechte Gestaltung des Dienstes und hierfür die Auswertung des Referrers tatsächlich zwingend erforderlich ist.[26] Dies wird im Folgenden bei der Prüfung des Erforderlichkeitsgrundsatzes erörtert.

Die vom Serviceprovider erstellten dokumentenbezogenen anonymen Nutzungsstatistiken dürfen nämlich ohne weitere Einschränkung letztendlich auch für andere Zwecke verwendet werden, für die eine Verarbeitung personenbezogener Daten nach § 15 Abs. 3 TMG ohne Einwilligung der Nutzer nicht zulässig wäre. Das gilt jedoch nur, wenn dieser andere Zweck mit den dazu bereits vorliegenden Statistikdaten erfüllt werden kann. Etwas anderes würde gelten, wenn die Dataprovider für den Mehrwertdienst andere Nutzungsprofile erstellen oder andere Auswertungen anhand der (personenbezogenen) Nutzungsdaten durchführen müssten und dieser Dienst weder unter Werbung noch Marktforschung oder bedarfsgerechter Gestaltung von Telemedien eingeordnet werden kann.

bb) Erforderlichkeitsgrundsatz

Nach dem Grundsatz der Erforderlichkeit dürfen auch für zulässige Zwecke nur diejenigen personenbezogenen Daten verarbeitet werden, die für den jeweiligen Zweck zwingend erforderlich sind. Es müssen also die von den Dataprovidern verarbeiteten Daten für den Zweck der bedarfsgerechten Gestaltung der Dokumenten- und Linkresolverserver (mithin für die Erzeugung aussagekräftiger Nutzungsstatistiken) erforderlich sein. Die Erforderlichkeit wurde jeweils wie folgt begründet:

- Die Angaben **IP-Adresse** und **Zeitpunkt der Ausführung** werden benötigt, um Zugriffe von gleichen Rechnern aus innerhalb der COUNTER-Doppelklickspanne zu erkennen und als lediglich einmaliger Download in der

[25] So soll zwar nach höchstrichterlicher Rechtsprechung Werbung jedes Verhalten sein, das darauf angelegt ist, andere dafür zu gewinnen, die Leistung desjenigen, für den geworben wird, in Anspruch zu nehmen (BGH, NJW 1992, 45).. Es ist aber zweifelhaft, ob dies so weit geht, dass auch Maßnahmen ganz generell zur Steigerung der Akzeptanz von Open-Access-Dokumenten bzw. des Open-Access-Publikationsmodells davon umfasst werden.
[26] Unabhängiges Landeszentrum für Datenschutz Schleswig-Holstein (ULD), Hinweise und Empfehlungen zur Analyse von Internet-Angeboten mit „Piwik“, S. 11, im Internet abrufbar unter: https://www.datenschutzzentrum.de/tracking/piwik/20110315-webanalyse-piwik.pdf

Statistik zu erfassen. Die **C-Klasse der IP-Adresse** wird darüber hinaus benötigt, um Crawler, also maschinelle Zugriffe nicht menschlichen Ursprungs zu erkennen und bei der Berechnung der Nutzungszahlen auszuklammern.

- Der **Dateipfad und -name**, die **Dokumenten-ID** und etwaige **Get-Parameter** sind erforderlich, um festzustellen, auf welches Dokument zugegriffen wurde. Im Webserverlogfile wird i. d. R. nur der Dokumentenpfad gespeichert. Anhand dieser Angabe ermittelt der Dataprovider dann, auf welches Dokument zugegriffen wurde und welche interne Dokumenten-ID dem Dokument zugewiesen ist. Zum Teil werden Dokumenten-IDs nach einem einheitlichen Standard gebildet (z. B. nach dem DOI-Standard – Digital Object Identifier), so dass der Service-Provider anhand dieser Angabe einrichtungsübergreifend feststellen kann, ob es sich bei zwei bei unterschiedlichen Data-Providern gespeicherten Dateien inhaltlich um dasselbe Dokument handelt.
- Die Angabe des **HTTP-Statuscodes** ist erforderlich, da nur erfolgreiche und gültige Anfragen in die Zählung eingehen sollen.
- Die Angaben **Größe des Dokuments in Byte** sowie die **Zahl der übertragenen Bytes** ist erforderlich, weil erst bei der Übertragung eines bestimmten Anteils des Dokuments (z. B. 95 %) ein Abruf als erfolgreicher Download gezählt werden soll.
- Die Angabe des **User-Agents** im HTTP-Header ist erforderlich, um Crawler zu identifizieren und deren Zugriffe bei der Erstellung der Nutzungszahlen auszuklammern. Webcrawler benutzen häufig individuelle User-Agent-Header (z. B. „*msnbot/1.0 (+http://search.msn.com/msnbot.htm)*“, anhand derer sie sich erkennen lassen.
- Für die im **Accept-Header** enthaltenen Angaben wie Dateitypen, Codierungen und Sprachen wurde keine weitere Erforderlichkeit benannt. Soweit diese Angaben also nicht erforderlich sind, dürfen sie auch nicht für die OA-Statistik in einem Webserverlogfile gespeichert und weiterverarbeitet werden.
- Die Angabe des **Referrers** ist erforderlich, um eine detaillierte Aufstellung geben zu können, von wo die Nutzer zu dem Dienst gelangt sind. Auf diese Weise kann der Einstieg optimiert werden. Wenn Nutzer in der Mehrzahl von Suchdiensten statt vom Portal des Anbieters kommen, muss die Sichtbarkeit des Dienstes auf dem Portal verbessert werden. Wenn umgekehrt keine Nutzer von Suchmaschinen kommen, muss eine Suchmaschinenoptimierung durchgeführt werden.

Die vorstehenden Begründungen für die Erforderlichkeit erscheinen vertretbar. Folgt man der Auffassung des ULD, dass die Speicherung und Auswertung des Referrers grundsätzlich auf Grundlage des § 15 Abs. 3 TMG für die Bildung von Nutzungsprofilen

zulässig ist, soweit dieser hierfür erforderlich ist, kann auch die Auswertung von Referrern noch gerechtfertigt werden.

Zwischenergebnis: Mit Ausnahme des Accept-Headers dürfen auf Grundlage des § 15 Abs. 3 TMG die o. g. Daten grundsätzlich für Zwecke der OA-Statistik in Webserverlogfiles gespeichert und von den Dataprovidern verwendet werden.

cc) Pseudonyme Nutzungsprofile

Dies ist jedoch nur unter weiteren Einschränkungen zulässig. Nach § 15 Abs. 3 TMG dürfen anhand der Nutzungsdaten nur „Nutzungsprofile bei Verwendung von Pseudonymen“ erstellt werden.

Der Begriff „**Nutzungsprofil**“ ist weder gesetzlich definiert noch finden sich Erläuterungen in der Gesetzesbegründung. In der Literatur wird unter Nutzungsprofil jede "systematische Zusammenstellung von Nutzungsdaten"[27] verstanden. An anderer Stelle ist gar davon die Rede, dass jeder Datensatz über eine Person, der zumindest ein Teilabbild über seine Persönlichkeit gibt, ein Nutzungsprofil darstellt bzw. dass ein Nutzungsprofil bereits bei Daten über einzelne Nutzungen ("Momentprofil") vorliegt.[28]

Daraus ergibt sich aber, dass als Nutzungsprofil nicht erst die Zusammenführung von Nutzungsdaten unterschiedlicher Dienste bzw. deren Zuordnung zu einer Person gilt, sondern dass es schon ausreichend ist, dass einzelne Datensätze vorliegen, die Auskunft über das Nutzerverhalten einer Person geben. Dies ist auch sachgerecht, da sich Diensteanbieter sonst den gesetzlichen Vorgaben dadurch entziehen könnten, dass sie Nutzungsdaten in „Rohform“ speichern, also weder zusammenführen noch systematisch auswerten und auf diese Weise entgegen § 13 Abs. 4 Nr. 2 TMG über das Nutzungsende hinaus speichern, aber weder pseudonymisieren noch den Nutzern ein Widerspruchsrecht einräumen.
Daher muss bereits ein einzelner Datensatz über einen Nutzungsvorgang als Nutzungsprofil angesehen werden. Ein solches Nutzungsprofil muss aber nicht auf eine einzelne Nutzung beschränkt sein. Zulässig sind daher auch Profile, die das Nutzungsverhalten innerhalb eines bestimmten Zeitraums umfassen.[29]

Ergänzend schreibt § 15 Abs. 3 TMG vor, dass die personenbezogenen Daten dabei durch **Pseudonyme** ersetzt werden müssen, d. h. durch Identifikationsmerkmale, die die Bestimmung der Betroffenen ausschließen oder wesentlich erschweren (vgl. § 3

[27] Bauer, MMR 2008, 435, 437.
[28] Schmitz, in: Spindler/Schmitz/Geis, TDG, § 6 TDDSG, Rn. 26
[29] Vgl. Schmitz, in: Spindler/Schmitz/Geis, TDG, § 6 TDDSG, Rn. 26.

Abs. 6 a BDSG[30]). Genauer gesagt müssen diejenigen Merkmale durch ein Pseudonym ersetzt werden, die einen Personenbezug ermöglichen, im vorliegenden Fall also die IP-Adressen der Nutzer. In diesem Sinn haben die obersten Aufsichtsbehörden für den Datenschutz im nicht-öffentlichen Bereich ausdrücklich festgestellt, dass IP-Adressen selbst keine Pseudonyme im Sinne des Telemediengesetzes sind.[31]

Durch eine Pseudonymisierung wird folglich bewirkt, dass zwar die verantwortliche Stelle (hier: der Dataprovider) die Nutzungsdaten noch einer bestimmten Person zuordnen kann, weil sie weiß oder zumindest wissen kann, welche Person sich hinter einem Pseudonym verbirgt; ein Dritter, der die Zuordnungsregel nicht kennt, kann dies aber nicht mehr bzw. nur noch mit unverhältnismäßig großem Aufwand an Zeit, Kosten und Arbeitskraft. Die Daten sind damit für den Dataprovider personenbezogen, für einen Dritten sind sie anonym. Mit der Pseudonymisierung wird außerdem bewirkt, dass auch die Dataprovider den Personenbezug während der weiteren Verarbeitung nicht mehr ohne Weiteres herstellen können, sondern es hierfür eines zusätzlichen Schrittes bedarf. Dementsprechend schreibt § 13 Abs. 4 Nr. 6 TMG auch vor, dass die Dataprovier durch organisatorische und technische Maßnahmen sicherstellen müssen, dass die Nutzungsprofile nicht mit den identifizierenden Angaben des Trägers des Pseudonyms zusammengeführt werden (dazu unten mehr).

Die Dokumenten- und Linkresolverserver speichern die Nutzungsdaten zumeist in Logfiles. Von dort sollen sie vom Dataprovider durch ein Programm ausgelesen und pseudonymisiert werden. Die **Zeitspanne zwischen der Erhebung der Nutzungsdaten und ihrer Pseudonymisierung ist so kurz wie möglich** zu halten, weil der Dataprovider grundsätzlich verpflichtet ist, die personenbezogenen Daten unmittelbar nach Beendigung des einzelnen Nutzungsvorgangs zu löschen (§ 13 Abs. 4 Nr. 2 TMG); **nach Ablauf des Nutzungsvorgangs dürfen sie folglich nur pseudonymisiert weiterverarbeitet werden**.[32]

Vertretbar dürfte es sein, dass der Dataprovider die Nutzungsdaten in Zyklen aus dem Logfile ausliest und die Logfiles in entsprechenden Zyklen (ggf. mit einer bestimmten Pufferzeit) überschrieben werden. Auf diese Weise werden die unveränderten Nutzungsdaten nur für den Zeitraum des Zyklus vorgehalten. Anhaltspunkte, wie lange dieser Zeitraum sein kann, finden sich in Literatur und Rechtsprechung, soweit ersichtlich, nicht. Allerdings bringt der Gesetzgeber durch die Wendung "unmittelbar nach de-

[30] Vgl. auf landesrechtlicher Ebene § 3 Abs. 7 LDSG.
[31] Datenschutzkonforme Ausgestaltung von Analyseverfahren zur Reichweitenmessung bei Internet-Angeboten, Beschluss der obersten Aufsichtsbehörden für den Datenschutz im nicht-öffentlichen Bereich am 26./27. November 2009, abrufbar unter: http://www.lfd.m-v.de/dschutz/beschlue/Analyse.pdf.
[32] Es sei denn eine Weiterverarbeitung in nicht pseudonymisierter Form ist aus anderen Gründen zulässig. Solche Gründe können z. B. Abrechnungszwecke sein, § 15 Abs. 2 TMG. Im Open-Access-Bereich dürfte dies jedoch keine Rolle spielen.

ren Beendigung" zum Ausdruck, dass er davon ausgeht, dass die Daten in sehr engem zeitlichen Zusammenhang nach der Beendigung des Nutzungsvorgangs gelöscht werden. Dieser Zeitraum dürfte sich im Bereich **weniger Minuten** bewegen. Eine längere Speicherung ohne Pseudonymisierung dürfte mit dem Gesetz unvereinbar sein.

Es ist beabsichtigt, die erhobene IP-Adresse beim Auslesen aus der Logdatei in den Dataprovider mittels einer Hashfunktion (SHA-256) unter Verwendung eines Salts in einen Hashwert umzuwandeln. Ein solcher Hashwert stellt eine Prüfsumme dar, die sich aus sich heraus nicht in den Ursprungswert zurückrechnen lässt. Eine IP-Adresse ergibt dabei stets denselben Hash-Wert. Damit die Ersetzung der IP-Adresse durch einen Hash-Wert den Anforderungen des TMG genügt, muss der Hash-Wert ein Pseudonym i. S. d. TMG sein. Das setzt voraus, dass der Hash-Wert einen Rückschluss auf die ursprüngliche IP-Adresse ausschließt oder zumindest wesentlich erschwert.
Ein solcher Hashwert wird nicht aufgrund einer geheimen, sondern vielmehr bekannten Zuordnungsregel gebildet (nämlich einen SHA – Secure Hash Algorithmus). Aus diesem Grund würde es nicht genügen, einen Hashwert lediglich aus der jeweiligen IP-Adresse zu berechnen. Derzeit werden noch überwiegend IP-Adressen des IPv4-Standards verwendet, ein Internetprotokoll, bei dem IP-Adressen eine Länge von 4 Oktetts haben. Mit IPv4 können rechnerisch maximal 2^{32} (= 4.294.967.296) IP-Adressen weltweit gebildet werden. Da jede IP-Adresse stets denselben Hash-Wert ergibt und die Rechenmethode allgemein bekannt ist, kann jede beliebige Person mit heute verfügbaren Rechnerkapazitäten sehr leicht und schnell aus allen weltweit existierenden IP-Adressen die zugehörigen Hash-Werte errechnen.[33] Für die rechtliche Bewertung spielt es keine Rolle, ob eine solche Reidentifikation beabsichtigt ist. Die Erstellung einer solchen Zuordnungsliste gesetzlich nicht verboten und wäre jedermann möglich. Daher würde auch eine vertragliche Verpflichtung der Projektpartner, von dieser Möglichkeit keinen Gebrauch zu machen, nichts an dem grundsätzlichen Personenbezug ändern.[34]

[33] Ist eine solche Zuordnung (Rainbow-Table) erst erstellt, kann ohne weiteres aus einem Hash-Wert auf die ursprüngliche IP-Adresse geschlossen werden. Dafür ist es nicht erforderlich, die Hash-Werte aller knapp 4,3 Milliarden IP-Adressen zu berechnen. Ca. 14 % des IP-Adressbereichs scheiden für Rechneranfragen an den Dokumentenserver sogar aus, weil es sich um Sonderbereiche gemäß „RFC 3330“ und „RFC 5735“ handelt (z. B. IP-Adressen 10.0.0.0 bis 10.255.255.255 reserviert für Netzwerke für den privaten Gebrauch). Sofern man nicht auf bestehende Rainbow-Tables zurückgreift, würde die Berechnung von 3,6 Milliarden IP-Adressen vollkommen ausreichend sein und ca. 11 Stunden dauern. Dies führt insgesamt dazu, dass jedermann ohne unverhältnismäßigen Aufwand an Zeit, Kosten und Arbeitskraft von einem Hash-Wert auf die ursprüngliche IP-Adresse schließen kann. Die vollständige Berechnung der IP-Hash-Paare der Universität Stuttgart dauert nur 3 Sekunden.

[34] An dieser Bewertung ändert auch § 13 Abs. 4 Nr. 6 TMG nichts, der es dem Diensteanbieter verbietet, die Nutzungsprofile mit Angaben zur Identifikation des Trägers des Pseudonyms zusammenzuführen. Denn letztlich würde die Ersetzung einer IP-Adresse durch einen Hashwert nicht zu einer Pseudonymisierung, sondern lediglich zur Ersetzung eines personenbezogenen Merkmals durch ein anderes führen. Die von § 15 Abs. 3 TMG geforderte Pseudonymisierung der Nutzerdaten wäre auf diese Weise nicht erfüllt.

Eine Verwendung der zulässigerweise erhobenen Nutzungsdaten ist deshalb nur zulässig, wenn die IP-Adresse durch ein Merkmal ersetzt wird, das eine Bestimmung der ursprünglichen IP-Adresse wesentlich erschwert. Da sich die Größe des IPv4-Adressraums nicht erweitern lässt, kann eine sinnvolle Pseudonymisierung nur dadurch geschehen, dass eine nicht allgemein bekannte, d. h. geheime Zuordnungsregel genutzt wird. Dies soll im Projekt durch die Verwendung eines Salts erfolgen. Hierfür wird an die IP-Adresse eine rechnerisch zufällig erzeugte Zeichenkette (das Salt) gefügt und aus beidem im SHA-Verfahren der Hashwert errechnet. Ein Rückschluss auf die ursprüngliche IP-Adresse ist dann nur demjenigen möglich, der das verwendete Salt kennt. Die Erzeugung von Zuordnungslisten von Hashwert zu IP-Adresse ist ohne Kenntnis vom Salt nicht möglich.[35] Das Salt muss mindestens 128 Bit bzw. 16 Zeichen lang sein und durch ein kryptografisches Verfahren erzeugt werden (z.B. Pseudozufallszahlengenerator mit komplexem Initialisierungsvektor). Außerdem muss das Salt geheim gehalten werden, darf also weder an den Serviceprovider noch an Dritte weitergegeben werden. Dies würde die Pseudonymisierung der Daten für den jeweiligen Empfänger aufheben. Aus diesem Grund muss das Salt auch (als entscheidendes Geheimnis für die Pseudonymisierung) in regelmäßigen Abständen geändert werden. Ein Zeitraum von vier Wochen bzw. von einem Monat dürfte ausreichend sein. Nach dem Wechsel des Salts ist das alte Salt zu löschen, weil es nicht mehr benötigt wird.
Dagegen ist es unschädlich, wenn sämtliche Dataprovider dasselbe Salt für die Pseudonymisierung verwenden und deshalb wissen, mithilfe welches Salts die anderen Dataprovider die IP-Adressen pseudonymisiert haben. Zwar sind die Nutzungsdaten eines Dataproviders damit auch für die übrigen Dataprovider nurmehr pseudonyme Daten (also nicht anonym), die **Dataprovider übermitteln untereinander jedoch weder die erhobenen Roh-Nutzungsdaten der Webserverlogfiles noch die die gehashten IP-Adressen enthaltenden Nutzungsdaten**. Sie erlangen daher keine Kenntnis von den bei den anderen Dataprovidern anfallenden Nutzungsdaten. Dies ist eine wesentliche Voraussetzung dafür, dass alle Dataprovider dasselbe Salt verwenden können.

Bei Einhaltung dieser Vorgaben werden damit im Ergebnis nurmehr pseudonyme Nutzungsprofile erstellt, wie von § 15 Abs. 3 TMG gefordert. Dabei ist – wie bereits angesprochen – zu beachten, dass die Dataprovider nach § 13 Abs. 4 Nr. 6 TMG verpflichtet sind, durch technische und organisatorische Maßnahmen sicherzustellen, dass die

[35] Ein Personenbezug kann sich theoretisch auch aus den weiteren Protokolldaten ergeben. Zu denken ist hier insbesondere an Zeitstempel. Da Zeitstempel oft auch an anderer Stelle im Netzwerk gespeichert werden (z. B. zur Störungserkennung und –beseitigung) und dort wiederum mit anderen personenbezogenen Daten verarbeitet werden (z. B. IP-Adresse), kann nicht ausgeschlossen werden, dass Dataprovider über den Zeitstempel die Protokolldaten für das Projekt OA-Statistik u. U. mit anderen Daten verknüpfen und so einen Personenbezug herstellen können. Da § 15 Abs. 3 TMG eine Anonymisierung der Nutzungsdaten durch den Diensteanbieter aber nicht verlangt und auch nur die verantwortliche Stelle über eine solche Verknüpfungsmöglichkeit verfügt, nicht aber Dritte, steht eine solche Verknüpfbarkeit mit anderen Daten der Erstellung von Nutzungsprofilen im vorgesehenen Verfahren nicht entgegen.

pseudonymen Nutzungsprofile nicht mit Angaben zur Identifikation des Trägers des Pseudonyms zusammengeführt werden können. Dies wird im vorliegenden Projekt allerdings bereits dadurch umgesetzt, dass die Nutzungsdaten in sehr kurzen zeitlichen Abständen von i. d. R. fünf Minuten an den Serviceprovider übertragen werden, der keine Kenntnis vom Salt hat und die Nutzungsdaten deshalb nicht auf die ursprüngliche IP-Adresse und damit auch nicht auf die einzelnen Nutzer zurückführen kann; **beim Dataprovider sind die Nutzungsdaten nach der erfolgreichen Übertragung zu löschen**, da sie nicht mehr benötigt werden.

dd) kein Widerspruch des Nutzers

Ein Nutzungsprofil darf jedoch nur erstellt werden, wenn ein Nutzer dem nicht widerspricht (so genanntes „Opt-Out"), § 15 Abs. 3 S. 1 TMG. Daraus folgt mehrerlei: zum einen haben die Dataprovider den Nutzern im Rahmen des Dokumentenserver- oder Linkresolverdienstes eine (effektive) Möglichkeit zum Widerspruch gegen die Erstellung von Nutzungsprofilen einzuräumen. Wie bereits ausgeführt entfällt dieses Recht nicht dadurch, dass die Dataprovider die Nutzungsdaten im Grunde nur im „Rohformat" verarbeiten, also keine systematischen Auswertungen oder Aggregationen vornehmen. Dem Zweck der Vorschrift entsprechend ist jegliche personenbezogene Erfassung des Nutzungsverhaltens als Nutzungsprofil anzusehen unabhängig davon, auf welchen Zeitraum sich das Profil erstreckt.

Zum anderen müssen die Dataprovider die Nutzer im Rahmen ihrer allgemeinen Unterrichtungspflicht nach § 13 Abs. 1 TMG (dazu unten mehr) **zu Beginn** der Nutzung des Dokumenten- oder Linkresolverservers auf dieses Widerspruchsrecht **hinweisen**, § 15 Abs. 3 S. 2 TMG.
Die Form des Widerspruchs ist gesetzlich nicht vorgegeben. Im Unterschied zur Einwilligung muss aber **der Nutzer aktiv handeln**, wenn er die Erstellung von Nutzungsprofilen **unterbinden** möchte. Daher ist es zulässig, den Webserver standardmäßig so zu konfigurieren, dass Nutzungsdaten gespeichert und für OA-Statistik in der oben dargestellten Weise verwendet werden und nur dann, wenn der Nutzer einen Widerspruch erklärt hat, eine solche Datenverwendung unterbleibt. **Eine aktive Zustimmung der Nutzer ist also für die Datenverarbeitung im Projekt OA-Statistik nicht erforderlich.**
Das Widerspruchsrecht darf jedoch nicht an Anforderungen geknüpft werden, die dieses Recht letztlich leerlaufen lassen. So wird es teilweise als nicht ausreichend angesehen, wenn die Nutzer bei nur vorübergehender Nutzung eines Dienstes per E-Mail (oder gar Fax) widersprechen müssen. Dies widerspreche der „interaktiven und schnellen Nutzungsform" im Internet.[36] Etwas anderes kann dann gelten, wenn der Betroffene

[36] Vgl. Schmitz, in: Spindler/Schmitz/Geis, TDG, § 6 TDDSG Rn. 28

einen Dienst langfristig nutzt, mit dem Anbieter etwa ein dauerhaftes Vertragsverhältnis eingeht, zu dessen Beginn er dann einmalig per E-Mail widersprechen muss. Dies wird bei den Repositorien- und Linkresolverdiensten aber selten der Fall sein. In der Regel wird eine nur eine kurzzeitige Nutzung stattfinden, bei der dem Nutzer dann eine entsprechend einfache Möglichkeit zum Widerspruch angeboten werden muss. Zu denken ist etwa an einen „Widerspruchsbutton“, eine Schaltfläche, auf die der Nutzer klicken kann, wenn er der Erstellung von Nutzungsprofilen widersprechen möchte. Technisch kann daran das Setzen eines temporären Cookies gekoppelt werden, der die Information enthält, dass der Nutzer der Erstellung von Profilen widersprochen hat. Dabei ist zu beachten, dass die Nutzer dann im Rahmen der allgemeinen Unterrichtungspflicht nach § 13 Abs. 1 TMG auf das Setzen des Cookies und die damit verbundene Datenverarbeitung hingewiesen werden müssen. Auf das Setzen eines dauerhaften Cookies sollte dagegen – auch angesichts der Tatsache, dass aufgrund europarechtlicher Vorgaben die Anforderungen an das Setzen vor allem dauerhafter Cookies verschärft werden sollen – verzichtet werden. Das Setzen eines dauerhaften Cookies kann demnach die Einwilligung der Nutzer erforderlich machen.

Eine andere Möglichkeit bestünde darin, die IP-Adresse des Nutzers, der den Button anklickt, in einer Liste zu speichern (Blacklist) und nach Ende des Nutzungsvorgangs alle Datensätze im Webserverlogfile zu verwerfen (zu löschen), die die entsprechende IP-Adresse enthalten. Der Blacklisteintrag wäre nach dem Ende des Nutzungsvorgangs ebenfalls zu löschen.

Widerspricht ein Nutzer, dürfen seine Nutzungsdaten **nicht für die Erstellung eines Nutzungsprofils verwendet werden**. Vielmehr müssen die angefallenen personenbezogenen Daten **unmittelbar nach dem Ende des Nutzungsvorgangs gelöscht** werden. Die Daten dürfen folglich keine Verwendung im Projekt OA-Statistik finden und zwar auch dann nicht, wenn sie hierfür „frühestmöglich“ pseudonymisiert werden.[37] Allenfalls zulässig wäre es, im Fall von Widersprüchen die Nutzungsdaten des Betroffenen in **anonymer Form** für das Projekt OA-Statistik zu verarbeiten.[38] Denn bei anonymer Verarbeitung sind die datenschutzrechtlichen Bestimmungen von vornherein nicht einschlägig, da es an einem Personenbezug bzw. einer Personenbeziehbarkeit fehlt.

[37] Sieht das Verfahren eine unverzügliche Löschung von Daten der widersprechenden Nutzer nicht vor und können diese auch nicht in der Protokolldatei ausgefiltert werden, müssen sämtliche Protokolldaten unverzüglich gelöscht werden, wenn ein Nutzer widerspricht, weil sie dann zumindest teilweise rechtswidrig verwendet werden.
Eine mangelnde Implementation einer Widerspruchsmöglichkeit birgt also die Gefahr, dass sämtliche Daten gelöscht werden müssen.
[38] Vgl. Schmitz, in: Spindler/Schmitz/Geis, TDG, § 6 TDDSG Rn. 31.

Exkurs: Anonymisierung

Für eine Anonymisierung müssen die Nutzungsdaten unmittelbar nach Beendigung der Nutzung anonymisiert werden, d. h. bereits im Webserverlogfile dürfen nur anonyme Daten gespeichert werden. Dies könnte zuverlässig dadurch geschehen, dass ausschließlich um die letzten beiden Byte gekürzte IP-Adressen des IPv4-Formats gespeichert werden.
Da in absehbarer Zukunft die flächendeckende Einführung von IPv6 bevorsteht und damit sowohl ein Paradigmenwechsel bei der Adressvergabe als auch Änderungen in den Anforderungen an die Anonymisierung der IP-Adresse einhergehen, sind aus unserer Sicht folgende Punkte zu beachten:

Im Gegensatz zu den IPv4-Adressen haben alle IPv6-Adressen eine festgelegte äußere Struktur. Dabei beschreiben die ersten 64 Bit der 128 Bit langen Adresse das Netzwerksegment und die letzten 64 Bit die Schnittstellen-ID, die einer Netzwerkschnittstelle in einem Endgerät zugeordnet ist. Im Unterschied zu IPv4-Adressen können aber sowohl das Netzwerksegment als auch die Schnittstellen-ID weltweit für sich allein eindeutig sein (→ Multihoming). Es kann aber auch sein, dass nur die Kombination aus Netzwerksegment und Schnittstellen-ID weltweit eindeutig ist. Die Eindeutigkeit bzw. Mehrdeutigkeit des Netzwerksegments ergibt sich einerseits aus der Länge des Präfixes (z.B. 32 Bit) und der möglichen, für den Netzwerkbetreiber bekannten, Strukturierung der übrigen Bits des Netzwerksegments (z.B. die zweiten 32 Bit). Die Eindeutigkeit der Schnittstellen-ID ergibt sich dann, wenn diese aus der MAC-Adresse der Netzwerkkarte generiert wird. Mehrdeutig ist sie bei aktivierten „Privacy Extensions", da die ID dann nur im Netzwerksegment eindeutig ist und sich regelmäßig ändert.

Wie auch schon bei den IPv4-Adressen, so ist auch bei den IPv6-Adressen eine Anonymisierung dahingehend zu unterscheiden, ob die Client-Adresse aus dem eigenen Netzwerk oder aus einem fremden stammt. Aber unabhängig hiervon ist die IPv6-Adresse zuerst immer um die Schnittstellen-ID, also die letzten 64 Bit, zu kürzen.

Je nach Art und Weise wie das Netzwerksegment aufgebaut ist, ist für den Netzwerkbetreiber der Teil der Bits, die nicht das Präfix ausmachen, einem Netzwerkteilnehmer eindeutig zuordenbar. Um welche Teile das Netzwerksegment in diesem Fall gekürzt werden muss, um eine Anonymisierung herzustellen, muss einer Einzelfallbetrachtung unterzogen werden und kann nicht pauschal beantwortet werden.

Für einen fremden Dritten ist das Netzwerksegment, ohne weitere Kenntnis, nicht grundsätzlich eindeutig zuordenbar. Er kann sich aber Kenntnis darüber beschaffen, welche Länge das Präfix hat, und kann dann die Annahme treffen, dass mit hoher

Wahrscheinlichkeit, bei hinreichend kurzem Präfix, eine eindeutige Zuordnung von Seiten des Netzwerkbetreibers besteht. Um hier eine eindeutige Anonymisierung zu gewährleisten, muss unseres Erachtens das Netzwerksegment zusätzlich um die letzten drei Oktetts gekürzt werden.

Um auf der sicheren Seite zu sein, müssen also bei einer IPv6-Adresse die letzten 88 Bit bzw. 11 Oktetts gekürzt werden. Die verbleibenden ersten 40 Bit bzw. 5 Oktetts bei einer IPv6-Adresse bieten die gleiche geografischen Genauigkeit bzw. gewollte Unschärfe, wie die um das letzte Oktett gekürzte IPv4-Adresse.

Aus Sinn und Zweck des Widerspruchsrechts ergibt sich, dass die Nutzer die Möglichkeit zum Widerspruch haben müssen, bevor Nutzungsprofile erstellt werden und damit bevor sie das inhaltliche Angebot des Webservers nutzen. Aus diesem Grund sollte ein entsprechender Button dem Nutzer bereits auf der Eingangsseite angezeigt werden.

e) Datenweitergabe an den zentralen Serviceprovider

Die von den Dataprovidern pseudonymisierten Daten sollen in regelmäßigen, kurzen Zeitabständen (i. d. R. fünf Minuten) an den zentralen Serviceprovider übertragen werden. Als Serviceprovider fungiert die VZG. Da die VZG lediglich die pseudonymisierten Nutzungsdaten erhält und dabei keine Kenntnis vom für die Pseudonymisierung verwendeten Salt erlangen soll, sind die **Nutzungsdaten für die VZG anonym**. Die VZG kann aus dem Hashwert der IP-Adresse mit verhältnismäßigem Aufwand an Zeit, Kosten und Arbeitskraft nicht mehr die ursprüngliche IP-Adresse bestimmen. Auch die weiteren vom Dataprovider zur Verfügung gestellten Daten kann sie nicht auf eine natürliche Person zurückführen, auch dann nicht, wenn sie die Datensätze aller Dataprovider zusammenführt. Anhand der Hashwerte kann der Serviceprovider lediglich feststellen, dass ein Nutzungsvorgang (wahrscheinlich) ein und demselben Nutzer zuzurechnen ist, jedoch nicht, wer dieser Nutzer ist.
Datenschutzrechtliche Vorschriften hat der Serviceprovider folglich beim weiteren Umgang mit diesen Daten nicht zu beachten, soweit die Datenverarbeitung bei und durch ihn erfolgt.
Etwas anderes würde dann gelten, wenn die einzelnen Datensätze mitsamt der Hashwerte der IP-Adressen an die Dataprovider zurückgemeldet werden würden, wenn dort das zur Berechnung des Hashwerts verwendete Salt noch gespeichert wäre. Der jeweilige Dataprovider könnte dann die Datensätze anhand der Hashwerte wieder den ursprünglichen IP-Adressen zuordnen. Ein solches Vorgehen ist jedoch nicht geplant.

Beabsichtigt ist vielmehr, dass der Serviceprovider anhand der übertragenen Daten dokumentenbezogene Nutzungszahlen erzeugt. Auf diese Weise werden aggregierte

Zugriffszahlen ermittelt, die selbst keinen Rückschluss mehr auf einzelne Nutzer zulassen. Auch diese Daten sind folglich anonym. Sie werden vom Serviceprovider an die Dataprovider zurückgemeldet.

4. Weitere Pflichten der Diensteanbieter

Die Dataprovider unterliegen als Diensteanbieter nach dem TMG bzw. als verantwortliche Stellen darüber hinaus den folgenden Pflichten.

a) Allgemeine Informationspflichten

Nach § 13 Abs. 1 TMG haben Diensteanbieter die Nutzer „zu Beginn des Nutzungsvorgangs“ zu unterrichten über

- die Art,
- den Umfang und
- die Zwecke

der Erhebung und Verwendung personenbezogener Daten sowie über etwaige Datenverarbeitungen außerhalb der EU. Die Betreiber von Repositorien und Linkresolverservern müssen die Nutzer daher z. B. in Form einer Datenschutzerklärung umfassend auch (aber nicht nur) über die Datenverarbeitung im Rahmen des Projekts OA-Statistik informieren. Darin ist auch über den Einsatz von Cookies aufzuklären (etwa zur technischen Realisierung des Widerspruchsrechts), weil auch hierfür personenbezogene Daten verarbeitet werden. Die Unterrichtung muss vollständig und auch für einen Durchschnittsnutzer verständlich sein. Die Hinweise oder ein Verweis auf diese muss so angebracht werden, dass ein Nutzer sie üblicherweise zur Kenntnis nimmt, wenn er das entsprechende Angebot aufruft. Die Information muss für den Nutzer jederzeit abrufbar sein. Es genügt daher nicht, die Hinweise einmalig beim Aufruf des Repositoriums oder des Linkresolverdienstes einzublenden. Der Nutzer muss die Erklärung vielmehr jederzeit (z. B. über einen Link) aufrufen können.

b) Verfahrensverzeichnis

Jede verantwortliche Stelle hat ein Verzeichnis zu führen, das Angaben über die automatisierten Verfahren enthält, mit denen personenbezogene Daten verarbeitet werden (so genanntes Verfahrensverzeichnis bzw. Verfahrensbeschreibung).[39] Eine solche automatisierte Datenverarbeitung erfolgt bei den Dataprovidern auch im Rahmen von OA-Statistik. Daher haben die Dataprovider die Datenverarbeitung, die sie als verant-

[39] Siehe z. B. § 11 LDSG, § 8 NDSG.

wortliche Stelle durchführen, in ein Verfahrensverzeichnis aufzunehmen. Welche Inhalte ein solches Verzeichnis haben muss, ergibt sich aus dem für die jeweilige Stelle einschlägigen allgemeinen Datenschutzgesetz (also für die Universität Stuttgart aus § 11 LDSG, für die Universität Göttingen aus § 8 NDSG u. s. w). Je nach Bundesland bestehen hier geringfügige Unterschiede. In einigen Bundesländern sind Formulare vom Verordnungsgeber vorgeschrieben, die verwendet werden müssen, in anderen Ländern stellen die Aufsichtsbehörden Muster zur Verfügung, die verwendet werden können.[40] In wieder anderen Fällen gibt es keine Vorlagen.
Bei der Erstellung eines solchen Verzeichnisses ist zunächst die Frage zu klären, ob die Datenverarbeitung für das Projekt OA-Statistik ein eigenständiges Verfahren darstellt oder ob sie lediglich Bestandteil anderer Verfahren ist. Hierfür kommt es auf den Einzelfall an, also zu welchen Zwecken der Webserver sonst noch Nutzungsdaten protokolliert und inwieweit diese sich mit den Zwecken der OA-Statistik überschneiden.

Verfahrensverzeichnisse enthalten einen öffentlichen Teil, soweit sie nicht ohnehin vollständig öffentlich sind. Die Angaben im öffentlichen Teil sind auf Antrag jedermann zugänglich zu machen.[41] Auskunft aus dem öffentlichen Teil kann daher nicht nur der betroffene Nutzer, sondern auch jeder beliebige Dritte verlangen. Im Übrigen können die Aufsichtsbehörden Einsicht in die Verfahrensverzeichnisse verlangen.

c) Auskunftsrecht

Grundsätzlich steht den Nutzern ein **Auskunftsrecht über die zu ihrer Person gespeicherten Daten** zu. Auf Verlangen ist ihnen mitzuteilen, welche Daten zu ihnen bzw. zu ihrem Pseudonym gespeichert sind (§ 13 Abs. 7 TMG i. V. m. § 34 BDSG). Eine solche Auskunftserteilung ist aber schon faktisch nur möglich, solange personenbezogene Daten gespeichert sind. Wenn – wie im Projekt OA-Statistik – die (pseudonymisierten) Nutzungsdaten nur wenige Minuten vom Webserver bzw. im Dataprovider gespeichert sind und unverzüglich nach der Übertragung an den Serviceprovider gelöscht werden, sind etwaige Auskunftsanfragen betroffener Nutzer in aller Regel schon wegen Zeitablaufs hinfällig; denn Auskunft zu geben ist nach § 34 BDSG nur über die aktuell zum Betroffenen gespeicherten Daten, nicht auch über inzwischen gelöschte Daten. Es ist deshalb davon auszugehen, dass die Speicherfrist der von den Dataprovidern im Projekt OA-Statistik erhobenen Daten zu kurz ist, als dass Auskunftsersuchen der Nutzer nach § 13 Abs. 7 i. V. m. § 34 BDSG erfolgreich beantwortet werden könnten.[42] Nutzer können auch nicht bei den Dataprovidern Auskunft über die beim

[40] So z. B. für das Land Niedersachsen: http://www.lfd.niedersachsen.de/download/32246
Saarland: http://www.datenschutz.saarland.de/images/stories/pdf/Datenschutzrecht/Verfahrensbeschreibung rtf
[41] § 8a NDSG, § 11 Abs. 7 LDSG, § 19 a Abs. 1 S. 5 BlnDSG, § 9 Abs. 2 SDSG.
[42] Etwas anderes kann dann gelten, wenn die Nutzungsdaten aus anderen Gründen (also außerhalb des Projekts OA-Statistik) länger gespeichert werden.

Serviceprovider gespeicherten Daten verlangen, da der Serviceprovider eine selbständige verantwortliche Stelle ist. Entsprechende Auskunftsersuchen wären daher direkt an den Serviceprovider zu richten.
Jedoch werden auch etwaige an den Serviceprovider gerichtete Auskunftsersuchen gegenstandslos sein. Beim Serviceprovider sind lediglich anonyme Daten gespeichert (s. o.). Über solche muss grundsätzlich keine Auskunft erteilt werden.

Daraus folgt jedoch nicht, dass die betroffenen Nutzer keinerlei Anspruch haben zu erfahren, inwieweit personenbezogene Daten im Rahmen des Projekts OA-Statistik verarbeitet werden. Ein Informationsanspruch ergibt sich – wie bereits dargestellt – zum einen aus der allgemeinen Unterrichtungspflicht der Diensteanbieter nach § 13 Abs. 1 TMG sowie aus dem Recht auf Einsichtnahme in den öffentlichen Teil des Verfahrensverzeichnisses (s. oben).

d) Sonstiges

Der Vollständigkeit halber ist darauf hinzuweisen, dass die Dataprovider bei der Gestaltung ihres Diensteangebots losgelöst von OA-Statistik die in § 13 Abs. 4 TMG genannten **technischen und organisatorischen Maßnahmen** umzusetzen haben. Sie müssen sicherstellen, dass

- der Nutzer die Nutzung des Dienstes jederzeit beenden kann,
- die anfallenden personenbezogenen Daten über den Ablauf des Zugriffs oder der sonstigen Nutzung unmittelbar nach deren Beendigung gelöscht oder ggf. gesperrt werden,
- der Nutzer die Telemedien gegen Kenntnisnahme Dritter geschützt in Anspruch nehmen kann,
- die personenbezogenen Daten über die Nutzung verschiedener Telemedien durch denselben Nutzer getrennt verwendet werden können,
- Daten nach § 15 Abs. 2 nur für Abrechungszwecke zusammengeführt werden können und
- Nutzungsprofile nach § 15 Abs. 3 nicht mit Angaben zur Identifikation des Trägers des Pseudonyms zusammengeführt werden können.

Im Übrigen sind die allgemeinen Datensicherheitsmaßnahmen zum Schutz von Webservern umzusetzen.

5. Zusammenfassung

Im Projekt OA-Statistik verarbeiten die Dataprovider personenbezogene Daten (Nutzungsdaten). Die Erhebung der Nutzungsdaten kann auf § 15 Abs. 1 TMG gestützt

werden. Die Speicherung der Nutzungsdaten in Webserverlogfiles und die weitere Verarbeitung der Daten beim Dataprovider ist ohne Einwilligung der Nutzer nur auf Grundlage von § 15 Abs. 3 TMG in pseudonymisierter Form zulässig; die Nutzer können gegen die Speicherung und weitere Verarbeitung ihrer Nutzungsdaten jedoch widersprechen. Unzulässig ist die Verarbeitung des Accept-Headers, soweit diese Daten für die Erzeugung von Nutzungsstatistiken nicht erforderlich sind.

Die Pseudonymisierung der IP-Adresse und damit auch der übrigen Nutzungsdaten muss innerhalb weniger Minuten nach der Erhebung der Daten mithilfe eines Salts erfolgen. Das Salt ist zwingend geheimzuhalten. Es darf dem Service-Provider nicht mitgeteilt werden. Da alle Dataprovider dasselbe Salt verwenden, dürfen sie untereinander keine Nutzungsdaten austauschen.

Die Dataprovider müssen weitere Anforderungen im Zusammenhang mit der Verarbeitung personenbezogener Daten beachten (allgemeine Hinweispflichten an die Nutzer, Erstellung eines Verfahrensverzeichnisses).

Die Nutzungsdaten sind für den Serviceprovider anonym. Der Serviceprovider hat deshalb keine datenschutzrechtlichen Anforderungen zu erfüllen. Die vom Serviceprovider erzeugten Nutzungsstatistiken sind ebenfalls anonym.

BS/MG/17.10.11