Technische Prüfung der Datenschutzerklärungen auf deutschen Hochschulwebseiten

Tim Wambach, Prof. Dr. Konstantin Knorr

Alle deutschen Hochschulen präsentieren sich im Internet über eine Webseite. Datenschutzaspekte spielen für Besucher dieser Seiten, wie zukünftige Studierende und Forschungspartner, eine immer wichtigere Rolle. Der vorliegende Beitrag untersucht die Existenz und den Inhalt der nach der deutschen Gesetzgebung verpflichtenden Datenschutzerklärungen (DSE). Ferner wird untersucht, ob sich die DSE mit den tatsächlichen Inhalten der Webseiten deckt. Die Methodik beinhaltet eine manuelle Überprüfung ausgewählter Webseiten und eine automatisierte Prüfung mittels eines modifizierten PyQt-Browsers mit dem Fokus auf der Erkennung von Trackern. Die Auswertung der Daten zeigt, dass viele DSE entweder fehlen, falsche Informationen beinhalten oder unvollständig sind. Der Artikel endet mit einer Diskussion der Ursachen und Empfehlungen zur Verbesserung der Missstände.

Zusätzliche Schlüsselwörter: Datenschutzerklärung, Hochschulen, Web-Tracking, Cookies, LDSG, BDSG, TMG

1 EINLEITUNG

Die Webseite ist das digitale Aushängeschild einer Hochschule. Studiengänge und Forschungsprojekte werden Interessenten darin präsentiert. Im Unterschied zu anderen Webseiten spielen dabei kommerzielle Interessen, zumindest bei den staatlichen Hochschulen, nur eine untergeordnete Rolle. Außerdem sind bei Hochschulen Mitarbeiter sehr frei bei der Gestaltung ihrer Webseiten und müssen häufig keinen zentralen Vorgaben entsprechen. Die Betreiber der Webseiten sind öffentliche Institutionen und keine Unternehmen. Datenschutzaspekte spielen bei Webseiten öffentlicher Betreiber eine immer wichtigere Rolle wie jüngst [Beltermann (2015a), Beltermann (2015b)] zeigten.

Selbstverständlich beinhaltet der Datenschutz an Hochschulen viele andere Fragestellungen [Witt (2004)], die allerdings interner Natur sind und nicht von externen Parteien untersucht werden können. Die Startseiten der Internetauftritte der Hochschulen sind frei zugänglich

Tim Wambach, M.Sc.
Universität Koblenz-Landau, Fakultät für Wirtschafts- und Verwaltungsinformatik, Projekt: Strukturwandel des Privaten, 56070 Koblenz, Deutschland,
https://www.uni-koblenz-landau.de/de/koblenz/fb4/iwvi/aggrimm/team-en/tim-wambach
E-Mail: wambach@uni-koblenz.de

Prof Dr. Konstantin Knorr
Hochschule Trier, Fachbereich Informatik, Am Schneidershof, Postfach 1826, 54208 Trier, Deutschland,
http://www.hochschule-trier.de/index.php?id=knorr
E-Mail: knorr@hochschule-trier.de

Die Erlaubnis zur Kopie in digitaler Form oder Papierform eines Teils oder aller Teile dieser Arbeit für persönlichen oder pädagogischen Gebrauch wird ohne Gebühr zur Verfügung gestellt. Voraussetzung ist, dass die Kopien nicht zum Profit oder kommerziellen Vorteil gemacht werden und diese Mitteilung auf der ersten Seite oder dem Einstiegsbild als vollständiges Zitat erscheint. Copyrights für Komponenten dieser Arbeit durch Andere als FHWS müssen beachtet werden. Die Wiederverwendung unter Namensnennung ist gestattet. Es andererseits zu kopieren, zu veröffentlichen, auf anderen Servern zu verteilen oder eine Komponente dieser Arbeit in anderen Werken zu verwenden, bedarf der vorherigen ausdrücklichen Erlaubnis.

und vermitteln dem Besucher ein Bild davon, welchen Stellenwert der Datenschutz bei der Hochschule genießt.

Wir haben uns daher in diesem Artikel das Ziel gesetzt, die Datenschutzerklärungen der deutschen Hochschulen manuell und automatisiert zu prüfen. Es wird nachgewiesen, dass nicht nur Benutzer bei der Interpretation einer DSE überfordert sind, sondern auch Hochschulen nicht in der Lage sind, ihrer Verpflichtung in ausreichender Weise nachzukommen.
Der Artikel gliedert sich dazu in folgende Abschnitte: Eine Einführung zu Datenschutz an den Hochschulen, Datenschutzerklärungen für Hochschulwebseiten und Web-Tracking gibt Abschnitt 2. In Abschnitt 3 wird eine manuelle Auswertung der Datenschutzerklärungen der zehn Hochschulen mit den meisten Studierenden durchgeführt. Abschnitt 4 prüft darauf aufbauend automatisiert die DSE von Hochschulen mit mehr als 2.000 eingeschriebenen Studenten. Beide Abschnitte beschreiben die dazugehörige Methode und die erhobenen Daten. Den Abschluss bildet Abschnitt 5 mit einer Diskussion der erhobenen Daten und verwandter Arbeiten sowie einem Ausblick auf zukünftige Betätigungsfelder.

2 GRUNDLAGEN

2.1 Datenschutz an deutschen Hochschulen

Die deutsche Hochschullandschaft umfasst aktuell 426 Hochschulen, die sich in die Gruppen staatlich, private und konfessionelle Hochschulen unterteilen lässt. Die älteste deutsche Hochschule ist die Ruprecht-Karls-Universität Heidelberg aus dem Jahr 1386. In den letzten Jahren wurde eine große Zahl neuer Hochschulen gegründet. Die Studierendenzahlen schwanken von weniger als 20 bis zu fast 80.000 an der Fernuniversität Hagen. In allen deutschen Bundesländern und großen Städten sind Hochschulen ansässig [Drachentoeter (2015)].
Die Hochschulen unterliegen den deutschen Datenschutzgesetzen. Die Hochschulen sind in der Regel Körperschaften des öffentlichen Rechts und zugleich staatliche Einrichtungen. Sie können aber auch in anderer Rechtsform errichtet werden. Die größte Gruppe der staatlichen Hochschulen unterliegt in der Regel dem Landesdatenschutzgesetz (LDSG) des entsprechenden Bundeslandes. Für andere Hochschulen kann auch das Bundesdatenschutzgesetz (BDSG) gelten. Laut den LDSG und BDSG sind die Hochschulen verpflichtet, einen Datenschutzbeauftragten (DSB) zu benennen. Dieser kann entweder hochschulintern benannt werden, oft z. B. ein Mitarbeiter des Rechenzentrums oder aus der Verwaltung (Rechtsabteilung) oder Professor. Extern kann z. B. eine Kanzlei oder ein anderes Unternehmen im Auftrag der Hochschule diese Aufgabe übernehmen. Eine zentrale Herausforderung des Datenschutzes ist, dass er (1) juristische und (2) technische Aspekte beinhaltet.
Die Hochschulen müssen per Datenschutzgesetz einen Datenschutzbeauftragen benennen. Der DSB übt eine komplexe und äußerst anspruchsvolle Tätigkeit aus, die technische, juristische und verwaltungsorganisatorische Fähigkeiten verlangt. Leider treten aufgrund der komplexen und sich häufig verändernden Hochschul-IT-Infrastruktur immer wieder „Datenpannen“ auf [Knoke (2015)].
Bei vielen Hochschulen ist ein Trend zu beobachten, den Datenschutz an Externe zu vergeben oder die Kompetenzen zu bündeln. Die Hochschulen in Baden-Württemberg unterhalten z. B. mit ZENDAS (Zentrale Datenschutzstelle der baden-württembergischen Universitäten) eine eigene Zentrale für den Datenschutz. In anderen Bundesländern gibt es regelmäßige Treffen der Hochschuldatenschützer, z. B. organisiert durch den Landesdatenschutzbeauftragen.
[Witt (2004)] gibt ein Überblick typischer Probleme beim Hochschuldatenschutz. [ZENDAS (2015)] liefert per wöchentlichem Newsletter aktuelle Informationen zum Hochschuldatenschutz.

2.2 Datenschutzerklärung für Hochschulwebseiten

Eine DSE ist laut Telemediengesetz (TMG) für alle Hochschulwebseiten verpflichtend. Nach § 13 TMG muss der Diensteanbieter den Nutzer zu Beginn des Nutzungsvorgangs über Art, Umfang und Zwecke der Erhebung und Verwendung personenbezogener Daten sowie über etwaige Weitergaben von Daten an Staaten außerhalb der EU unterrichten. Nach § 15 Abs. 3 TMG darf der Diensteanbieter für Zwecke der Werbung, der Marktforschung oder zur bedarfsgerechten Gestaltung der Telemedien Nutzungsprofile bei Verwendung von Pseudonymen erstellen, sofern der Nutzer dem nicht widerspricht. Der Dienstanbieter hat den Nutzer auf sein Widerspruchsrecht im Rahmen der Unterrichtung nach § 13 Abs. 1 hinzuweisen.
Der Aufbau und die Zuständigkeit für die Erstellung und Pflege der DSE sind bei den Hochschulen sehr unterschiedlich geregelt. Bei der Erstellung sind typischerweise Parteien wie das Rechenzentrum, Web-Redakteure, der Datenschutzbeauftragte und der Kanzler bzw. die Rechtsabteilung involviert. Oft wird die DSE als Teil des Impressums publiziert. Für die Struktur und den Aufbau der DSE gibt es keine klaren Vorgaben. Neben den rechtlich bindenden Vorgaben gibt es Aspekte, die z. B. von der OECD [OECD (2015)] empfohlen werden.

2.3 Web-Tracking

Web-Tracking im Internet hat in den vergangenen Jahren einen deutlichen Zuwachs erfahren. [Schneider et al. (2014)] analysieren den Einsatz von Web-Tracking Verfahren quantitativ. Damit Benutzer verlässlich zwischen verschiedenen Webangeboten verfolgt werden können, ist aus technischer Sicht die Einbindung eines externen Objekts, wie eines Scripts, Tracking-Pixel, Web-Bugs oder ähnliches notwendig. Durch Aufruf einer Webseite, die eine solche Einbindung vorgenommen hat (Embedder), wird eine Übermittlung an einen Tracking- oder Werbedienst (Tracker) bewirkt. Das Interesse des Webseitenbetreibers für den Einsatz von Web-Trackern kann betriebswirtschaftlich, z. B. zur Verbesserung der Marketingstrategie, begründet sein. Zudem stehen umfangreiche Analysetools kostenfrei zur Verfügung (Google Analytics). Für den Benutzer bringt eine solche unbemerkte Zusammenführung von Daten, die aus Webaktivitäten entstanden sind, jedoch nicht nur Vorteile.
In [Schneider et al. (2014)] wird gezeigt, dass Doubleclick und Adition die häufigsten Tracker sind, die auf deutschen Webseiten gesichtet wurden (Google Analytics wurde in dieser Arbeit nicht betrachtet). In [W3Tech (2015)] zeigt sich, dass ca. 50 % der Webseiten (weltweit) durch Google Analytics analysiert werden. Trackingmethoden dieser Art können allerdings vom Benutzer bemerkt werden. Es existieren bereits diverse Browsererweiterungen zum Selbstdatenschutz, z. B. Ghostery [Ghostery (2015)], die aber eine automatisierte Auswertung in ihren AGBs untersagen. Neben diesen technischen Maßnahmen werden Besucher durch Gesetze wie Landes-, Bundesdatenschutz- oder Telemediengesetz geschützt, sofern das Webangebot unter deutsches Recht fällt. Besonders zu erwähnen sind die „Impressumspflicht“ (§ 5 TMG) und Angaben zum Umgang mit personenbezogenen Daten (§ 13 TMG) in einer sog. DSE.

Aktuell ist es mit Standard-Browsern und deren Browsererweiterungen nur schwer möglich, Werbe- oder Trackingverfahren auf Webseiten verlässlich zu detektieren und zu messen. Dies ist nicht zuletzt der Vielfalt der Trackingmethoden geschuldet, die in [Mayer und John (2012)] näher beschrieben werden. Viele Cross-Domain-Tracking-Verfahren bewirken jedoch eine Verbindung vom Browser zum Trackinganbieter. Allein ein solcher Verbindungsaufbau ist bereits mit der Übergabe der IP-Adresse zu einer bestimmten Uhrzeit verbunden. Darüber hinaus wird durch die Verwendung von Cookies eine Zuordnung über mehrere Anfragen hinweg erleichtert. Zusätzliche Verbindungen, die nicht dem angeforderten Webauftritt angehören, werden im Folgenden als externe Verbindungen bezeichnet. Wird vom Benutzer bspw. eine Webseite angefordert, die Google Analytics zur Erhebung und Auswertung von Besucherstatistiken verwendet, baut der Browser im Hintergrund eine Verbindung zum Webserver von Google Analytics auf.

3 AUSWERTUNG DER DATENSCHUTZERKLÄRUNGEN DER ZEHN GRÖSSTEN HOCHSCHULEN

Im Juli 2015 wurden die Webseiten der zehn deutschen Hochschulen mit den meisten Studierenden (nach Angaben in [Drachentoeter (2015)]) untersucht. Zunächst wurde nach der DSE auf der Hauptseite gesucht. Falls diese gefunden wurde, wurde deren Inhalt geprüft und nach folgenden Kriterien aus den Kategorien (1) BASIS, (2) INHALT und (3) OECD bewertet. Tabelle 1 gibt eine Übersicht der Ergebnisse der Untersuchung.

	Basis				Inhalt				OECD			
Hochschule	Anzahl Wörter in DSE	Version	DSE vorhanden	Verweise auf Gesetze	HTTP-Logs	Web Tracker	Social Media	Werbung	Zweckbestimmung	Sicherung	Mitspracherecht	Ansprechpartner
Fernuniversität in Hagen	382	26.08.2014	IMP	LDSG	Ja	PIWIK	K.A.	K.A.	Ja	Nein	Nein	Ja-1
Ludwig-Maximilians-Universität München	2293	K.A.	Ja	LDSG, TMG	Ja	PIWIK	Facebook, Twitter	K.A.	Ja	Nein	Nein	Ja-2
Universität zu Köln	893	K.A.	Ja	K.A.	Ja	K.A.	Facebook, Google, Twitter, Youtube	Adition	Ja	Nein	Nein	Ja-1
Johann Wolfgang Goethe-Universität Frankfurt am Main		K.A.	FEHLT	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.
Ruhr-Universität Bochum	138	22.07.2015	IMP	K.A.	K.A.	PIWIK	Facebook	K.A.	Ja	Nein	Nein	Ja-1
Westfälische Wilhelms-Universität (Münster)	0	K.A.	FEHLT	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.	K.A.
RWTH Aachen	424	14.05.2014	Ja	LDSG, TMG	Ja	K.A.	K.A.	K.A.	Ja	Nein	Ja	Ja-2
Universität Duisburg-Essen		07.04.2015	FEHLT	TMG	Ja	K.A.	Facebook, Twitter, Google+	K.A.	Ja	Nein	Nein	Ja-1
Universität Hamburg	614	K.A.	Ja	K.A.	K.A.	Google Analytics	Facebook	K.A.	Nein	Nein	Ja	Ja-2
Friedrich-Alexander-Universität Erlangen-Nürnberg		K.A.	IMP	K.A.	K.A.	K.A.	K.A.	K.A.	Nein	Nein	Nein	Ja-1

Tab. 1: Ergebnisse der manuellen Prüfung der DSE der größten deutschen Hochschulwebseiten (K.A. = „Keine Angaben“) (Quelle: Eigene Darstellung)

Die Kategorie BASIS umfasst:

- Die Spalte „Anzahl Wörter in DSE“ gibt an, wie viele Wörter die DSE enthält. Natürlich hängt die Länge der DSE von der Komplexität der spezifischen Hochschulinfrastruktur ab. Trotzdem gibt die Wortanzahl einen groben Anhaltspunkt, welche Bedeutung der Datenschutz auf der Webseite genießt.
- Version: Ist die DSE mit einem Datum oder einer Versionsnummer versehen? Das Fehlen einer Version erschwert es, bei wiederholten Besuchen Änderungen an der DSE zu erkennen.
- DSE vorhanden: Existiert eine separate DSE (=> „Ja“), ist sie Teil des Impressums (=>„IMP“) oder war sie nicht auffindbar („FEHLT“)? Laut einem aktuellen Urteil des OLG Hamburg sollte die DSE separat vom Impressum gehalten werden [Overbeck (2014)].
- Verweise auf Gesetze: Wird in der DSE auf einschlägige Gesetze wie das jeweilige LDSG, BDSG oder TMG verwiesen? Damit wird einem Besucher verdeutlicht, auf welcher Rechtsgrundlage die DSE steht.

Die Kategorie INHALT klärt, ob die DSE die technischen Aspekte HTTP-Logs, Web-Tracker, Social Media und Werbung adressiert und benennt ggf. die verwendeten Technologien.
Innerhalb der Kategorie OECD werden die folgenden Aspekte untersucht, die aus den OECD-Grundsätzen [OECD (2015)] abgeleitet sind:

- Zweckbestimmung: Wird der Grund für die Erhebung von personenbezogenen (PBZ) Daten angegeben?
- Sicherung: Werden Sicherheitsvorkehrungen genannt, mit denen die PBZ-Daten geschützt sind?
- Mitspracherecht: Wird der Benutzer auf die ihm nach BDSG zustehenden Rechte wie Auskunftsrecht und Löschrecht hingewiesen?
- Ansprechpartner: Wird ein Ansprechpartner für Datenschutzfragen genannt? Mögliche Antworten: „Nein", „Ja-1": Allgemeine Kontaktdaten, „Ja-2": Datenschutzspezifische Kontaktdaten

Nur vier von zehn Hochschulen haben eine separate DSE, drei haben gar keine, bei drei Hochschulen ist die DSE ins Impressum integriert. Die Länge schwankt zwischen 138 (Bochum) und 2.293 Wörtern (München). Nur bei vier von zehn ist die verwendete Version erkennbar. Vier von zehn Seiten nennen die rechtlichen Grundlagen. Fünf von zehn Seiten protokollieren die HTTP-Zugriffe. PIWIK (3) und Google Analytics (1) sind die erwähnten Tracker. Facebook (5) und Twitter (3) sind die am häufigsten erwähnten Social-Media-Seiten. Die Uni Köln nutzt als einzige Werbung (Adition).

Den Zweck der eigenen Datenerhebung nennen die meisten Seiten (=> statistische Auswertung, Abwehr/Erkennung von Angriffen). Die Verwendung von Werbung oder Social Media wird i.d.R. nicht begründet. Keine Seite nennt getroffene Sicherheitsvorkehrungen. Das Mitspracherecht wird nur bei zwei Seiten erwähnt. Drei Seiten haben spezifische Ansprechpartner für den Datenschutz, fünf allgemeine, zwei gar keinen.

4 AUTOMATISIERTE TRACKERERKENNUNG AUF HOCHSCHULWEBSEITEN

4.1 Methodik

4.1.1 Technische Grundlagen

Die Erhebung externer Verbindungen wäre z. B. durch Überwachung der Netzwerkkommunikation mittels tcpdump oder Wireshark möglich. Diese Methode hätte allerdings den Nachteil, dass unklar ist, an welcher Stelle die Webseite vollständig geladen wurde und die Überwachung beendet werden kann. Ebenfalls sind fehlerhafte Ergebnisse aufgrund der Aktivitäten anderer Anwendungen (im Hintergrund) nicht ausgeschlossen. Wir haben daher einen Browser so modifiziert, dass die erstellten externen Verbindungen nach dem Aufruf einer Webseite ausgelesen werden können. Die Python-Bibliothek PyQt [PyQt (2015)] beinhaltet einen Browser, der zur Verarbeitung gängiger aktueller Techniken (HTML, XHTML, CSS, JavaScript, HTML canvas, AJAX) in der Lage ist. Zur Analyse wurde das Framework durch die Einbindung einer eigenen QNetworkAccessManager-Klasse abgeleitet und so modifiziert, dass jede Verbindung durch den Browser protokolliert wird und anschließend ausgewertet werden kann. Die durchgeführten Änderungen werden in Listing 1 genauer dargestellt.

```
class NetworkAccessManager(QNetworkAccessManager):
  requests = []
  [...]
  def createRequest(self, operation, request, data):
    self.requests.append([ request.url().encodedHost(),
                           request.url().encodedPath()])

    return QNetworkAccessManager.createRequest(self,
                                               operation,
                                               request,
                                               data)
  def getRequests(self):
      return self.requests
```

Listing 1: Modifikation der QNetworkAccessManager-Klasse (Quelle: Eigene Darstellung)

Sobald alle Bestandteile der untersuchten Webseite vollständig geladen sind, ist der Analysevorgang abgeschlossen und gibt das Ergebnis aus, nämlich eine Auflistung sämtlicher HTTP-Anfragen (getRequests). Zur näheren Verdeutlichung ein Beispiel: Durch Abruf der Webseite http://www.uni-hamburg.de wurden am Tag der Erhebung, neben den Verbindungen zum Webserver der Universität selbst, auch Verbindungen zu serveby.flashtalking.com, t4ft.de, google-analytics.com und imagesrv.adition.com registriert. Vom Framework wird der HTTP User-Agent „User-Agent: Mozilla/5.0 (Windows NT 6.1; Win64; x64) AppleWebKit/534.34 (KHTML, like Gecko) python Safari/534.34“ verwendet, ein durchaus typischer Eintrag, der auf aktuellen Windows-Plattformen Verwendung findet.

Die Überprüfung von Webseiten auf solche externen Verbindungen soll im Folgenden auf eine Testmenge angewendet werden, von der ein gewisses Maß an Sorgfalt im Umgang mit dieser Thematik erwartet wird. Dabei handelt es sich um Webauftritte von deutschen Hochschulen, die gemäß [Drachentoeter (2015)] mehr als 2.000 eingeschriebene Studierende (Stand: Juli 2015) aufweisen. Dies trifft auf 207 Einrichtungen zu. Bei dieser Analyseform wird lediglich die Hauptseite ausgewertet, d. h. es werden Verbindungen ausgewertet, die durch Aufrufen der Startseite erzeugt werden. Die Unterseiten bleiben von der Analyse unberührt. Automatische Weiterleitungen werden ausgeführt und berücksichtigt. Tag der Erhebung ist der 19.7.2015.

4.1.2 Überprüfung der Datenschutzerklärung

Nachdem nun eine Technik zur Erkennung von externen Verbindungen umgesetzt wurde, stellen wir die Frage, ob diese innerhalb der DSE berücksichtigt werden. Insbesondere beim Einsatz von Tracking-Verfahren muss dem Benutzer eine Widerspruchsmöglichkeit eingeräumt werden (§ 15 Abs. 3 TMG). Hierfür ist es notwendig, das Impressum bzw. die DSE des Webauftritts zu suchen.

Die Suche wird auf den Hauptseiten der Hochschuleinrichtungen durchgeführt. Im ersten Schritt wird auf der Hauptseite nach Links zu „Impressum“, „Datenschutz“, und „Privacy“ gesucht. Sofern eine solche Verlinkung gefunden werden kann, handelt es sich dabei um das gesuchte Dokument. Falls nur ein Link zum Impressum gefunden wird, wird anschließend in diesem nach Erwähnungen oder einer Verlinkung zu „Datenschutz“ oder „Privacy“ gesucht.

Werden im Impressum selbst Aussagen zum Datenschutz getroffen, sind Impressum und DSE zusammengefasst.
Auf diese Weise findet eine automatische Überprüfung statt, ob Impressum und DSE vorhanden sind. In den Fällen, in denen diese Dokumente nicht auffindbar waren, wurde anschließend eine manuelle Suche durchgeführt und die Daten ergänzt. Hierbei zeigten sich für einen Besucher möglicherweise unerwartet beiläufige Erwähnungen zum Datenschutz in sog. „Disclaimern", in denen ebenfalls Aussagen zum Urheberrecht der Inhalte getroffen wurden. Ebenfalls zeigten sich Fälle, bei denen Impressum und DSE nicht direkt auf der Hauptseite gefunden werden konnten, sondern eine Suche benötigten, bspw. http://www.unibw.de/.
Zur Auswertung von DSE muss der derzeitige IST-Zustand eines Webauftritts bestimmt und anschließend mit dem SOLL-Zustand aus der DSE verglichen werden. Nachdem nun eine Technik zur Bestimmung von externen Verbindungen entwickelt und eine Auflistung der Datenschutzerklärungen der Hochschulwebseiten zur Verfügung steht, kann nun ein automatischer Abgleich erfolgen. Hierfür muss zunächst festgelegt werden, nach welchen Begriffen innerhalb der DSE gesucht werden soll. Aus den beobachteten externen Verbindungen wurden die häufigsten mit einem Bezug zum Web-Tracking ausgewählt und genauer auf ihre Nennung in der DSE überprüft (vgl. Tabelle 2).
Zur automatisierten Analyse werden, sofern eine Verbindung zu einem dieser Dienste registriert wurde, Impressum und DSE nach einer Erwähnung des Dienstnamens selbst und dem Namen des Unternehmens durchsucht. Wird bspw. eine Verbindung zu google-analytics.com auf der Hauptseite detektiert, wird in den erwähnten Dokumenten sowohl nach „Google Analytics", als auch nach „Google" gesucht. Findet eine Erwähnung statt, wurde diese Weitergabe bei Erstellung der DSE bedacht. Andernfalls handelt es sich aufgrund der fehlenden Widerspruchsbelehrung (§ 15 Abs. 3 TMG) um eine nicht vollständige DSE, vgl. Tabelle 4.

4.2 Ergebnisse

4.2.1 Ergebnisse der Untersuchung von externen Verbindungen

In diesem Abschnitt werden die Ergebnisse der Untersuchung nach externen Verbindungen zusammengefasst. Die meisten Einbindungen standen in Verbindung zu Google, weshalb dieser ein eigener Abschnitt gewidmet ist (vgl. Abschnitt 4.2.2).
In 105 von 207 Fällen wurden keine externen Verbindungen registriert. In Tabelle 2 sind die Serveradressen von Ressourcen zu sehen, die in den verbleibenden 102 Fällen am häufigsten innerhalb der Hochschulwebseiten eingebunden wurden (fett markiert) sowie die Anzahl der Hochschulen, die eine solche Einbindung unternommen haben.

HOSTNAME	ANZAHL	GOOGLE-SERVICE	Bemerkungen
google-analytics.com	29	Ja	
ajax.googleapis.com	14	Ja	
fonts.googleapis.com	11	Ja	
www.googleadservices.com	10	Ja	
www.google.com	9	Ja	
code.jquery.com	8	Nein	JS-Bibliothek
imagesrv.adition.com	8	Nein	Werbeanbieter
doubleclick.net	7	Nein	Trackingdienst
s.ytimg.com	6	Ja	YouTube Vorschaubilder
www.youtube.com	6	Ja	
googletagmanager.com	4	Ja	
googleapis.com	3	Ja	
maps.google.com, translate.google.com , ssl.google-analytics.com, translate.google.com, translate.googleapis.com	2	Ja	
google.de, googletagservices.com, oauth.googleusercontent.com, tcp.googlesyndication.com	1	Ja	

Tab. 2: Top 10 (fett markiert) der am häufigsten registrierten externen Verbindungen & Google-Services (Quelle: Eigene Darstellung)

4.2.2 Externe Verbindungen zu Google

In den Analysen zeigte sich, dass 61 der 207 Hochschulwebseiten eine Verbindung zu einem Google-Service bewirkten. Hierbei wurde der Google-Service anhand der Zeichenkette „google“ im Hostnamen bestimmt. Services, die ebenfalls direkt oder indirekt Google angehören, wurden in diesem Fall nicht berücksichtigt, so z. B. „doubleclick.com“, das im Jahre 2007 von Google übernommen wurde. Der Grund hierfür ist, dass nicht in allen Fällen eine vollständige Erfassung geschäftlicher Beziehungen zwischen verschiedenen Tracking-Unternehmen möglich ist und deshalb zunächst nur die offensichtlichen Verbindungen betrachtet werden. In Tabelle 2 sind die zusätzlichen Google-Services aufgeführt und in Abbildung 1 ist der daraus resultierende Graph dargestellt.

4.2.3 Verwendung von Google Analytics

Die Testmenge wurde auf Einbindungen von google-analytics.com überprüft. Eine solche Einbindung fand in 29 Fällen statt. Google Analytics erlaubt eine Anonymisierung der Benutzer durchzuführen. In [Google (2015)] ist beschrieben, wie das letzte Oktett der IP-Adresse auf den Servern von Google entfernt wird – bei IPv6-Adressen die letzten 80 Bit. Ob diese Funktion genutzt wird, ist im Quelltext der Webseite oder über den ‚aip‘-Parameter der HTTP-Anfrage erkennbar. Die Kürzung wird jedoch erst nach der vollständigen Übertragung von Google selbst durchgeführt. Eine Übersicht der ‚aip‘-Verwendung bei den Hochschulen ist in Tabelle 3 einsehbar.

HOCHSCHULE	AIP
Macromedia Hochschule für Medien und Kommunikation	Ja
Hochschule für nachhaltige Entwicklung Eberswalde	Nein
Steinbeis-Hochschule Berlin	Ja
SRH Hochschule Heidelberg	Nein
Hochschule für Wirtschaft und Umwelt Nürtingen-Geislingen	Nein
AKAD Bildungsgesellschaft (Stuttgart)	Nein
Universität Hohenheim (Stuttgart)	Ja
Hochschule für angewandtes Management (Erding)	Ja
Hochschule für angewandte Wissenschaften Coburg	Ja
Universität Bayreuth	Ja
Diploma Hochschule (Bad Sooden-Allendorf)	Ja
Wilhelm Büchner Hochschule (Pfungstadt)	Ja
Europäische Fernhochschule Hamburg	Ja
HFH Hamburger Fern-Hochschule	Ja
Universität Hamburg	Ja
Private Fachhochschule Göttingen	Ja
HAWK Hochschule Hildesheim/Holzminden/Göttingen	Nein
Leuphana Universität Lüneburg	Ja
Hochschule Hamm-Lippstadt	Nein
Rheinische Fachhochschule Köln	Nein
FOM Hochschule (Essen)	Ja
Fachhochschule Kaiserslautern	Ja
Universität Koblenz-Landau	Ja
Hochschule Mittweida	Ja
Hochschule für Technik, Wirtschaft und Kultur Leipzig	Nein
Hochschule Anhalt (Bernburg, Dessau und Köthen)	Ja
Fachhochschule Nordhausen	Ja
Fachhochschule Schmalkalden	Ja
Ernst-Abbe-Fachhochschule Jena	Ja

Tab. 3: Nutzung von Google-Analytics auf Hochschulwebseiten (Quelle: Eigene Darstellung)

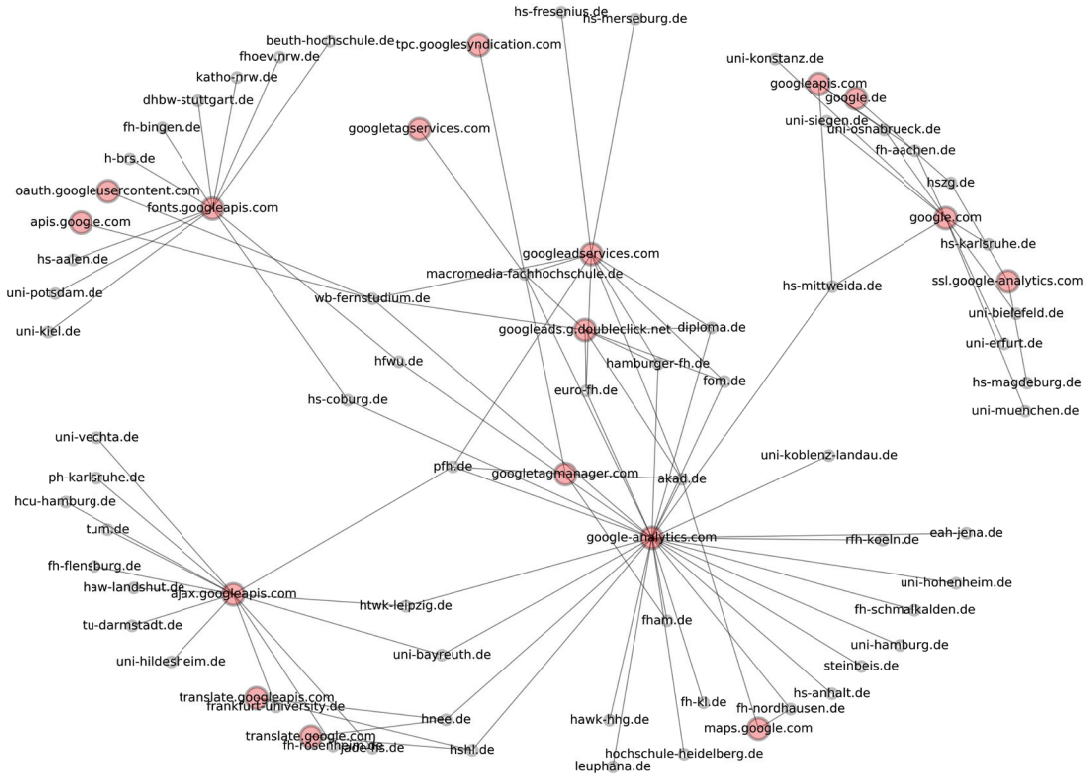

Abb. 1: Verbindungen von Hochschulwebseiten zu Google-Services (Quelle: Eigene Darstellung)

4.2.4 Ergebnisse der Untersuchung von Impressum und Datenschutzerklärung

In 20 der 207 Fälle waren keine expliziten Angaben zum Datenschutz auffindbar. In 137 Fällen wurde die DSE mit dem Impressum zusammengefasst bzw. war unter der gleichen Adresse abrufbar.

In 40 Fällen wurden Verbindungen zu den vier Anbietern (1) Google Analytics, (2) Doubleclick, (3) Adition oder (4) Facebook auf der Hauptseite detektiert. In 30 dieser Fälle wurden diese zumindest in Impressum oder DSE erwähnt. In den zehn verbleibenden Fällen war die DSE nachweislich unvollständig. Das Ergebnis dieser Analyse ist in Tabelle 4 einsehbar, wobei aus Gründen der Übersichtlichkeit die Negativ-Fälle gelistet sind. Es wurden zunächst nur diese vier Anbieter geprüft, da bei diesen klarer Bezug zu Tracking, Werbung und/oder Social Media erkennbar ist. Weitere Einbindungen wie z. B. „code.jquery.com“ wurden nicht weiter verfolgt.

Obwohl in vielen Datenschutzerklärungen explizit Facebook erwähnt wurde, konnte nur auf zwei Hauptseiten (http://diploma.de, https://www.akad.de) eine Verbindung zu Facebook gefunden werden. Dies ist möglicherweise der fehlenden Analyse der Unterseiten geschuldet; weitere Gründe werden in Abschnitt 5 behandelt.

HOCHSCHULE	google-analytics	doubleclick	adition	facebook
Universität der Künste Berlin			N	
Technische Universität Berlin			N	
Hochschule Karlsruhe Technik und Wirtschaft			N	
Hochschule Fresenius (Idstein)		N		
Universität Hamburg	J		N	
Rheinische Friedrich-Wilhelms-Universität Bonn			N	
Rheinische Fachhochschule Köln	N			
Universität Siegen			N	
Universität Koblenz-Landau	J		N	
Hochschule Magdeburg-Stendal	N			

Tab. 4: Ergebnis der automatischen Überprüfung von Datenschutzerklärungen.
„J": Erwähnung des Trackers in der Datenschutzerklärung,
„N": keine Erwähnung,
„Leeres Feld": keine externe Verbindung registriert.
(Quelle: Eigene Darstellung)

5 DISKUSSION UND AUSBLICK

5.1 Diskussion

Die Auswertung der manuellen Prüfung zeigt, dass drei der deutschen Hochschulen mit den meisten Studierenden gar keine DSE aufweisen und nur vier eine separate DSE besitzen. Die im TMG geforderten Punkte sind in den vorhandenen DSE weitgehend berücksichtigt. Sinnvolle Erweiterung wie z. B. eine Erwähnung der getroffenen Sicherheitsvorkehrungen oder des Mitspracherechts finden keine oder nur selten Erwähnung. Auffällig ist auch der starke Einsatz von Trackern, Social Media und vereinzelt auch Werbung. Leider wird dies in der DSE i.d.R. nicht begründet.

Die automatisierten Auswertungen zeigen, wie stark auf den 207 Hochschulwebseiten in Deutschland die Einbindung von externen Ressourcen genutzt wird. In ~50 % der Fälle fand eine Einbindung einer externen Ressource statt und 30 % der Hochschulen nutzen einen Dienst von Google. 14 % der deutschen Hochschulwebseiten nutzen zur Analyse Google Analytics, wobei in acht Fällen auf die Anonymisierungsoption zum Schutz der Besucher verzichtet wurde. Auf 20 Hochschulwebseiten (~10 %) war keine DSE auffindbar und es konnte nachgewiesen werden, dass diese in mindestens zehn Fällen (~5 %) unvollständig war.

In mindestens 30 Fällen (~15 %) besteht aus diesem Grund ein akuter Handlungsbedarf, eine klare DSE zu erarbeiten oder bestehende Fehler auszubessern. Ebenfalls ist bedenklich, dass auf fast jeder dritten Hochschulwebseite eine Übermittlung der IP-Adresse an Google stattfindet. Hier sollte dringend geprüft werden, ob Einbindungen von Google APIs, Karten, Übersetzungs- oder sonstigen Diensten tatsächlich notwendig sind.

Die Verwendung von Werbediensten, wie Adition, ist auf Webauftritten, insbesondere bei staatlich finanzierten Hochschulen, nur schwer nachvollziehbar und sollte von den Betreibern überdacht werden. Wie Tabelle 4 zeigt, wird der Einsatz von Adition deutlich häufiger verschwiegen als der von Facebook, was u. a. an der stärkeren „Sichtbarkeit" von Facebook liegen kann. Offensichtlich bewerten viele Hochschulen den monetären Anreiz beim Einsatz von Adition höher als den Datenschutz. Die Auswertung zeigte darüber hinaus, dass auf keiner Hauptseite eine Einbindung von Facebook vorgenommen wurde, ohne dies innerhalb der DSE zu erwähnen, obwohl in ca. 40 % der Hochschulwebseiten (Impressum, DSE) eine Aussage zu Facebook enthalten ist. Ein möglicher Grund hierfür könnte das starke Medieninteresse sein, das bei

anderen Trackingdiensten weniger präsent ist. Insbesondere in [ULD (2015a), ULD (2015b)] zeigt sich, wie intensiv das Thema Facebook von Aufsichtsbehörden verfolgt wird. So ist einerseits möglich, dass Facebook vorsichtshalber in die DSE aufgenommen wurde, oder dass die Einbettungen im Laufe der Zeit entfernt wurden.
Die manuelle Analyse einer DSE erlaubt eine deutlich tiefere inhaltliche Prüfung, z. B. die Prüfung der OECD-Kriterien in Tabelle 1, einer kleinen Anzahl von Seiten und ergänzt daher die automatisierte Untersuchung einer großen Zahl von Seiten. Beim Vergleich der Ergebnisse zeigt sich, dass viele Datenschutzerklärungen externe Verbindungen nur unvollständig auflisten oder externe Inhalte ankündigen, die auf den von uns untersuchten Seiten nicht verwendet wurden. Dass Analysewerkzeuge nicht immer auf der Hauptseite eingesetzt werden, zeigte sich z. B. bei http://www.uni-giessen.de/, bei der PIWIK erst auf einer der Unterseiten aktiv wird.
Die Ursachen der identifizierten Missstände sehen wir in folgenden Punkten: (1) der sich durch Gerichtsurteile ständig ändernde rechtliche Rahmen für die DSE, (2) Komplexität und Dynamik der Hochschul-IT, (3) oft unklare Zuständigkeit bei der Erstellung und Pflege der DSE. In allen Fällen könnte eine Hochschul-Muster-DSE helfen (vgl. Abschnitt 5.3). Ein weiteres Problem ist der Umfang des Webauftritts einer Hochschule, der häufig mehrere tausend Seiten umfasst. Die DSE soll möglichst für viele dieser Seiten gelten, was leicht zu einer Über- bzw. Unterdeckung führt.

5.2 Verwandte Arbeiten

In [Wambach (2015)] werden über ein Sandbox-Verfahren detailliertere Untersuchungen von Trackingverfahren auf Webseiten durchgeführt. In [Beltermann (2015a)] ist eine Untersuchung von 35 Behördenwebseiten zu finden, wovon neun Tracking-Verfahren einsetzen, ohne eine Widerspruchsmöglichkeit zu bieten. In [Beltermann (2015b)] können diesbezüglich vereinzelte Stellungnahmen der Behörden betrachtet werden. Dabei wurde insbesondere das Fehlen einer Widerspruchsmöglichkeit innerhalb der DSE gemäß § 15 Abs. 3 TMG von den Webseiten-Betreibern des Bundesverwaltungsgerichtes bestätigt.
Die Platform for Privacy Preferences (P3P) [P3P (2015)] bietet eine formelle Sprache, mit der spezifiziert werden kann, welche Daten von einem Webserver gespeichert werden, wozu diese Daten verwendet werden und wie lange sie gespeichert werden. Dies würde eine automatisierte Auswertung der DSE erheblich vereinfachen. Leider wird P3P kaum eingesetzt, da es von vielen potentiellen Nutzern als zu kompliziert angesehen wird und viele gängige Browser P3P nicht unterstützen.
Die strukturierte Analyse von DSE ist in letzter Zeit auch bei medizinischen Android-Apps durchgeführt worden [Sunyaev et al. (2014), Knorr et al. (2015)].

5.3 Zukünftige Arbeiten

Der letzte Abschnitt dieses Beitrags widmet sich zukünftigen Arbeiten, die die vorliegende Arbeit ergänzen und erweitern. Die vorgestellte Methodik kann z. B. um folgende Aspekte erweitert werden: (1) Berücksichtigung weiterer Technologien wie z. B. Cookies, (2) Untersuchung von anderen Klassen von Webseiten wie z. B. Seiten der öffentlichen Verwaltung, (3) zusätzliche Untersuchung von Unterseiten. Unterseiten können z. B. über sog. Spider wie Scrapy (scrapy.org) in einem ersten Schritt erfasst und dann mittels unseres modifizierten Browsers untersucht werden.
Die Hochschulen könnten stark von einer Hochschul-Muster-DSE profitieren, in der die häufigsten Anwendungen und Inhalte technisch und juristisch aufbereitet enthalten sind. Dieses Muster müsste dann nur noch auf die Hochschulspezifika angepasst bzw. erweitert werden.
Da in dieser Arbeit nur die technische Sichtweise dargestellt wird, wäre auch eine juristische Interpretation notwendig. So ist z. B. offen, ob Verbindungen zu Diensten wie code.jquery.com in der DSE benannt werden müssen.

DANK

Die Arbeit von Tim Wambach wurde durch die VolkswagenStiftung im Projekt „Strukturwandel des Privaten“ gefördert.

REFERENZEN

Beltermann, E. (2015a): Benutzerverfolgung durch staatliche Websites, https://netzpolitik.org/2015/benutzerverfolgung-durch-staatliche-websites/, zuletzt aufgerufen am 29. Juli 2015.

Beltermann, E. (2015b): Benutzerverfolgung durch staatliche Websites: Die Antworten, https://netzpolitik.org/2015/benutzerverfolgung-durch-staatliche-websites-die-antworten/, zuletzt aufgerufen am 29. Juli 2015.

Drachentoeter (2015): Liste der Hochschulen in Deutschland, https://de.wikipedia.org/wiki/Liste_der_Hochschulen_in_Deutschland, zuletzt aufgerufen am 20. Mai 2015.

Ghostery (2015): Add-ons, https://addons.mozilla.org/de/firefox/addon/ghostery/, zuletzt aufgerufen am 19. Juli 2015.

Google (2015): IP Anonymization in Google Analytics, https://support.google.com/analytics/answer/2763052?hl=en, zuletzt aufgerufen am 30. Juli 2015.

Knoke, F. (2015): Datenschützer an Unis: „Wir sind zahnlose Papiertiger", http://www.spiegel.de/unispiegel/studium/datenschuetzer-an-unis-wir-sind-zahnlose-papiertiger-a-585232.html, zuletzt aufgerufen am 19. Juli 2015.

Knorr, K., Aspinall, D. und Wolters, M. (2015): On the Privacy, Security and Safety of Blood Pressure and Diabetes Apps, in: Proc. of IFIP SEC 2015 International Conference on ICT Systems Security and Privacy Protection, May 26-28, Hamburg, 2015, S. 571-584.

Mayer, R. J. und John, C. M. (2012): Third-Party Web Tracking: Policy and Technology, in: Proceedings of the 2012 IEEE Symposium on Security and Privacy, 2012, S.12.

OECD (2015): OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, http://www.oecd.org/sti/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldata.htm, zuletzt aufgerufen am 6. August 2015.

Overbeck, A. (2014): Trenn oder Zahl!, in: DFN-Infobrief, 04, 2014, S. 5-7.

P3P (2015): Platform for Privacy Preferences (P3P) Project, Enabling smarter Privacy Tools for the Web, http://www.w3.org/P3P/, zuletzt aufgerufen am 1. August 2015.

PyQt (2015): PyQt Whitepaper, http://www.riverbankcomputing.com/static/Docs/PyQt4/pyqt-whitepaper-a4.pdf, zuletzt aufgerufen am 7. Juli 2015

Schneider, M., Enzmann, M. und Stopczynski, M. (2014): Web-Tracking-Report 2014, Darmstadt, 2014.

Sunyaev, A., Dehling, T., Taylor, P.L. und Mandl, K.D. (2014): Availability and quality of mobile health app privacy policies, in: Journal of the American Medical Informatics Association, 2014, S. 28-33.

ULD (2015a): Unabhängiges Landeszentrum für Datenschutz, https://www.datenschutzzentrum.de/facebook/, zuletzt aufgerufen am 27. Juli 2015

ULD (2015b): Unabhängiges Landeszentrum für Datenschutz, https://www.datenschutz-hamburg.de/ihr-recht-auf-datenschutz/internet/facebook.html, zuletzt aufgerufen am 25. Juli 2015

Wambach, T. (2015): Dynamische Trackererkennung im Web durch Sandbox-Verfahren, in: Tagungsband der DACH Security Konferenz 2015.

W3Tech (2015): Web Technology Surveys: Usage of traffic analysis tools for websites, http://w3techs.com/technologies/overview/traffic_analysis/all, zuletzt aufgerufen am 3. August 2015.

Witt, B.C. (2004): Datenschutz an Hochschulen, Ein Praxishandbuch für Deutschland am Beispiel der Universitäten Baden-Württembergs, Ulm, 2004.

ZENDAS (2015): Newsletter Verwaltung, https://www.zendas.de/zendas/newsletter_verwaltung/, zuletzt aufgerufen am 19. Juli 2015.