DGRI e. V. • Emmy-Noether-Str. 17 • D-76131 Karlsruhe

Bundesverfassungsgericht
- Erster Senat -
Postfach 1771

76006 Karlsruhe

per Fax vorab: 07 21 – 91 01 382

Dr. Anselm Brandi-Dohrn, maître en droit
1. Vorsitzender
Rechtsanwalt
Oranienstraße 164, D-10969 Berlin

Telefon: +49-30-61 68 94 09
Telefax: +49-30-61 68 94 56
E-Mail: abrandi-dohrn@boetticher.com

Berlin, 28. April 2014

In der Verfassungsbeschwerde des Herrn T.
AZ: 1 BvR 16/13

Hier: Stellungnahme der Deutschen Gesellschaft für Recht und Informatik e.V. nach § 27a BVerfGG

Sehr geehrter Herr Vizepräsident,

wir beziehen uns auf die Anfrage des Ersten Senats im obigen Verfahren vom 5. Dezember 2013 und danken für die der **Deutschen Gesellschaft für Recht und Informatik e.V.** (DGRI) gemäß § 27a BVerfGG gewährte Gelegenheit zur Stellungnahme.

Die DGRI ist eine der in Deutschland führenden unabhängigen wissenschaftlichen Vereinigungen im Bereich des IT-Rechts. Zu ihren Mitgliedern zählen Richter, Rechtsanwälte, Rechtswissenschaftler, Firmenjuristen der IT-Branche und IT-Techniker. Sie befasst sich mit Fragen im Bereich der Schnittstelle zwischen Informatik- und EDV-Recht einerseits sowie Recht und Wirtschaft andererseits. Sie fördert die Zusammenarbeit von Lehre, Forschung, Gesetzgebung und Praxis in allen Fragen der Informationstechnik. Sie begleitet Gesetzgebungsvorhaben als neutrale Institution und ist nicht den Partikularinteressen einzelner Unternehmen oder Branchen verpflichtet.

Der Vorstand der Gesellschaft hat sich nach schriftlichen Vorarbeiten auf seiner Sitzung vom 21./22. Februar 2014 mit dem Gegenstand der Verfassungsbeschwerde und insbesondere den vom Ersten Senat aufgeworfenen Fragen befasst,

- *inwieweit und auf welchen Wegen es Internetportalen wie dem Spiegel-Online-Archiv möglich ist, Einfluss auf die von Suchmaschinen aufgefundenen und ausgeworfenen Ergebnisse zu nehmen, und*
- *auf welche Weise und mit welchem Aufwand nachträglich die Erreichbarkeit personenbezogener Daten erschwert oder – im online-Zugriff – verhindert werden kann.*

Zu diesen Fragen nimmt die Gesellschaft wie folgt Stellung:

A. Vorbemerkung

Die folgenden Ausführungen befassen sich mit der Frage, wie der Zugriff auf bestimmte Informationen erschwert bzw. verhindert werden kann, die über ein bestimmtes Online-Portal zugänglich sind. Sie befassen sich nicht mit der Problematik, dass diese Informationen durch andere Online-Angebote kopiert und parallel auf Dritt-Portalen zugänglich gemacht werden können. Allerdings können die nachfolgenden Ausführungen auch auf derartige Kopien angewandt werden, doch wird sich in der Praxis eher das Problem ergeben, die Verantwortlichen für solche Kopien zu identifizieren und zu belangen.

B. Möglichkeiten von Internet-Portalen, auf die von Suchmaschinen aufgefundenen und angezeigten Ergebnisse Einfluss zu nehmen

Internet-Portalen stehen grundsätzlich drei Möglichkeiten zur Verfügung, um darauf Einfluss zu nehmen, dass bestimmte Inhalte nicht in den Ergebnislisten von Internet-Suchmaschinen angezeigt werden: Sie können

- Inhalte von ihren Webseiten entfernen oder ändern und so mittelbar auch eine Indexierung der betreffenden Inhalte durch Suchmaschinen verhindern bzw. aufheben;
- durch bestimmte Programmier-Befehle die Suchmaschinen anweisen, diese Inhalte nicht in ihren Such-Index aufzunehmen oder
- den Suchmaschinen bestimmte Inhalte gar nicht erst bereitstellen.

1. Technischer Hintergrund der Funktionsweise von Suchmaschinen

Suchmaschinen sind Computerprogramme, die Datenbanken nach bestimmten Informationen anhand festgelegter, geheimer Algorithmen durchsuchen und dem Suchenden die Ergebnisse mittels einer Trefferliste präsentieren, hierbei werden die Ergebnisse nach dem Grad der Übereinstimmung mit der Suchanfrage angeordnet.[1] Die heutigen Internet-Suchmaschinen funktionieren im Wesentlichen als so genannte *Crawler*[2] – Computerprogramme, die ausgehend von ihrem bereits bekannten Bestand an Internet-Adressen (URLs) versuchen, automatisch einen möglichst großen Teil des Internet-Angebots zu erfassen.[3] Dazu ruft der Crawler eine ihm bekannte URL ab, wertet deren Inhalt für seine Datenbank aus und nimmt insbesondere auch sämtliche in dem abgerufenen Angebot enthaltenen Links auf andere Angebote in seinen Bestand an URLs auf. Ausgehend von der Annahme, dass jede Datei im Internet mit anderen Dateien verlinkt ist, sollte sich mit diesem Vorgehen ein in der Theorie vollständiges Abbild des Internets erstellen lassen.

Suchmaschinen beschränken sich dabei üblicherweise nicht nur auf HTML-Dateien normaler WWW-Seiten, sondern beziehen auch andere Dateiformate wie Bilder, Text-Dateien oder – im vorliegenden Fall besonders relevant – PDF-Dateien in ihre Suche mit ein. Soweit PDF-Dateien Text als Schriftzeichen enthalten, können diese Inhalte ohne größeren Aufwand wie HTML-Dateien automatisiert von

[1] Vgl. *Egermann*, in: Kilian/Heussen, Computerrecht, 32. Ergänzungslieferung 2013, Suchmaschinen, Rn. 1.
[2] Auch Spider oder Agent genannt – vgl. *Spieker*, MMR 2005, 727.
[3] Vgl. zur Funktionsweise von Suchmaschinen: *Milstein/Lippold*, NVwZ 2013, 182, 183.

Web-Crawlern durchsucht und markiert werden. Gerade bei der Digitalisierung älterer Medien werden die Dokumente aber typischerweise gescannt und liegen lediglich als Bild vor, welches wiederum in das PDF-Dokument eingebettet ist. Um die so gescannten Dokumente für die digitalen Archive der Medien zu erschließen, werden die Bild-Dateien üblicherweise per Texterkennung[4] zurück in Text umgewandelt. Oftmals stellen die Medien sowohl den Text in Form einer HTML-Seite als auch das gescannte Original als PDF-Datei in ihr Online-Archiv ein.[5] Die Texterkennung aus gescannten Texten wird allerdings auch von Suchmaschinen genutzt. So ist beispielsweise von der am häufigsten verwendeten Suchmaschine[6] Google bekannt, dass das Unternehmen in PDF-Dokumenten enthaltene Bilder intern in Text umwandelt und diesen für die Auswertung der Web-Seite verwendet. Obwohl also bestimmte (PDF-)Dokumente technisch gesehen keine Textzeichen enthalten, kann ihr Inhalt mit einem Crawler ausgelesen werden.

Auf eine Suchanfrage hin wirft die Suchmaschine die nach einem Algorithmus[7] gewichteten, bei den Crawler-Durchläufen gefundenen Ergebnisse aus. Dabei werden nicht nur Ergebnisse mit dem direkten Suchwort ausgegeben, sondern seit längerem gehört es schon zum technischen Standard, dass auch Flexionen des eigentlichen Suchworts sowie einfache Synonyme angezeigt werden.

Darüber hinaus sind Suchmaschinen auch in der Lage, unterschiedliche Internetseiten zum gleichen Thema eigenständig zu verknüpfen und zu präsentieren, obwohl kein unmittelbarer syntaktischer Zusammenhang auf den Seiten existiert.

Google etwa wertet die URL der Seite und andere öffentlich verfügbare Informationen wie z. B. den Ankertext in Links zu der Website oder Angaben zu der Seite im Web-Verzeichnis Open Directory Project aus[8]. Anders als Suchmaschinen wie „Bing“ oder „ixquick“, die – soweit ohne eingehende Prüfung ersichtlich – nur Seiten aufführen, die tatsächlich die konkreten Suchbegriffe beinhalten, führt Google eine weitere Liste von Zusammenhängen in die Suchergebnisse ein.[9]

2. Entfernung bzw. Anonymisierung von Inhalten

Eine mittelbare Einflussnahme auf die von Suchmaschinen aufgefundenen und ausgeworfenen Ergebnisse erreicht ein Webseitenbetreiber zunächst dadurch, dass (i) entweder die betroffenen Beiträge vollständig oder teilweise aus dem Online-Archiv entfernt oder (ii) zumindest die Namen der Betroffenen anonymisiert werden. Damit würde nicht nur eine ursprüngliche Indizierung der betreffenden Inhalte, also des gesamten Beitrages oder nur des Namens des Betroffenen, verhindert, sondern auch nachträglich eine schon erfolgte Indizierung beeinflusst. Denn die Indizierung durch Suchmaschinen ist kein einmaliger Prozess, sondern wird in ständig wiederkehrenden Intervallen wiederholt. Wird ein Inhalt einer Webseite geändert, etwa dadurch, dass er ganz oder teilweise entfernt wird, wird diese Änderung bei der nächsten intervallmäßigen Indizierung berücksichtigt[10] und

---

[4] *OCR, Optical Character Recognition.*
[5] So verfahren beispielsweise der „Spiegel“ und das „Hamburger Abendblatt“.
[6] Vgl. http://www.seo-united.de/suchmaschinen.html.
[7] Die von den Suchmaschinenbetreibern verwendeten Algorithmen sind meist deren Geschäftsgeheimnis und daher nicht öffentlich bekannt – vgl. *Egermann*, in: Kilian/Heussen, Computerrecht, 32. Ergänzungslieferung 2013, Suchmaschinen, Rn. 1.
[8] https://support.google.com/webmasters/answer/156449?hl=de.
[9] Beispielsweise führt die Google-Suche nach dem Begriff „Paul T.“ auch zu Suchergebnissen, auf deren Seiten der konkrete Suchbegriff überhaupt nicht erscheint. So wird beispielsweise an vierter Stelle (Abrufergebnisse vom 14.02.2014) der Wikipedia-Eintrag „Apollonia (Kriminalfall)“ aufgeführt, der den Namen nicht enthält. Diese Listung liegt auch nicht etwa daran, dass der Name „Paul T.“ in früheren Versionen dieses Wikipedia-Eintrags vorhanden war, sondern lässt sich damit erklären, dass der Wikipedia-Eintrag einen Link auf einen Spiegel-Artikel beinhaltet, der den Namen „Paul T.“ enthält.
[10] Bei PDF-Dokumenten vorausgesetzt, dieses wurde nach Änderung mit neuem Dateinamen abgespeichert, vgl. unten C. 2. e).

werden Suchergebnisse in Bezug auf den entfernten Inhalt nicht mehr angezeigt, soweit sich der Bezug zum entfernten Inhalt nicht – wie im vorstehenden Abschnitt beschrieben – auch aus anderen Informationen ergibt, etwa Verlinkungen.

3. Anweisungen an Suchmaschinen

Die übliche Methode, bestimmte Web-Inhalte vor Suchmaschinen zu „verstecken", wird dadurch bewerkstelligt, dass auf der Webseite für die Crawler eine Anweisung vorgehalten wird, bestimmte Inhalte nicht aufzurufen bzw. nicht in den Index aufzunehmen.[11]

Dafür besteht seit 1994 der Quasi-Standard „*Robots Exclusion Protocol (REP)*".[12] Zu einer tatsächlichen Standardisierung - im Internet üblicherweise als „Request for Comments" (RFC) der Internet Engineering Task Force (IETF) und der Internet Society (ISOC) - ist es nicht gekommen. Die wichtigsten REP-Anweisungen werden aber von allen bedeutenden Suchmaschinen, allen voran Google und Bing, beachtet.[13] Allerdings existieren auch Ergänzungen zu den Standardbefehlen, die nur von einzelnen Suchmaschinen beachtet werden. Die REP-Anweisungen können auf verschiedene Arten einem Crawler mitgeteilt werden:

a) "/robots.txt" zum Ausschluss des Abrufs durch Crawler

Die ursprünglich einzige Methode zur Nutzung des REP bestand darin, im Wurzelverzeichnis (*root directory*) der Webseite eine Datei namens „robots.txt" anzulegen.[14] In dieser Textdatei werden dann für alle oder einzelne Suchmaschinen Angaben gemacht, welche Teile des Angebots (nicht) besucht werden sollen.[15] Ein „Disallow"-Eintrag in der robots.txt weist die Suchmaschinen an, die entsprechenden Dateien überhaupt nicht zu indizieren. Ein Beispiel für ein vollständiges „Verbot" für Suchmaschinen auf der kompletten Webseite würde lauten:

```
User-agent: *
Disallow: /
```

Soll dagegen nur eine bestimmte Seite, etwa nur die Seite http://www.dgri.de/68/Stellungnahmen.htm nur von einem ganz bestimmten Crawler, z.B. nur dem Googlebot, nicht abgerufen werden, während alle anderen Crawler alles abrufen dürfen, könnte die robots.txt lauten:

```
User-agent: Googlebot
Disallow: /68/Stellungnahmen.htm

User-agent: *
Disallow:
```

Soll das gesamte Unterverzeichnis „/archiv" nicht indiziert werden – etwa weil dort alte Print-Ausgaben des Spiegel im Volltext veröffentlicht werden, die in einer Form identifizierend berichten, die heute unzulässig wäre –, könnte die robots.txt wie folgt aussehen:

---

[11] Vgl. *Sieber*, in: Hoeren/Sieber/Holznagel, Multimedia-Recht, 36. Ergänzungslieferung 2013, Teil 1 Technische Grundlagen, Rn. 103, der robots.txt sowie Meta-Tags aufführt.
[12] Vgl. http://www.robotstxt.org/orig.html.
[13] Vgl. *Sieber*, in: Hoeren/Sieber/Holznagel, Multimedia-Recht, 36. Ergänzungslieferung 2013, Teil 1 Technische Grundlagen, Rn. 103.
[14] Bsp.: http://www.dgri.de/robots.txt; allgemeine Informationen und Programmierungsvorgaben für /robots.txt : vgl. http://www.robotstxt.org/robotstxt.html.
[15] Vgl. https://support.google.com/webmasters/answer/156449?hl=de.

```
User-agent: *
Disallow: /archiv/
```

Damit würden sämtliche Dateien, die im Verzeichnis „/archiv“ oder dessen Unterverzeichnissen liegen, nicht mehr von den Crawlern der Suchmaschinen aufgerufen.

Allerdings führt ein Disallow-Eintrag in der robots.txt technisch nicht dazu, dass die betroffenen Inhalte unter keinen Umständen von Suchmaschinen aufgeführt werden. Soweit mit einem Link von anderen – nicht „verbotenen“ und daher indizierten – Webseiten auf die gesperrten Dateien verlinkt wird, speichert die Suchmaschine auch diesen Link. Nach eigener Darstellung greift zB Google zumindest auf den Dateinamen (URL) und die Informationen des Web-Verzeichnisses Open Directory Project zu, sowie auf den Text, der auf der anderen Webseite im Zusammenhang mit dem Link publiziert ist.[16] In diesem Fall erscheint in der Darstellung der Suchergebnisse zur jeweiligen URL allerdings ein Hinweis darauf, dass dieser Inhalt nicht zugänglich ist – was den User aber nicht hindert, diese URL anzuklicken und so auf die (eigentlich gesperrte) Webseite zu gelangen. Somit besteht durch die Verwendung von robots.txt zwar ein Schutz dagegen, dass die Crawler seriöser Suchmaschinen die gesperrte Seite abrufen, nicht aber dagegen, dass die gesperrten Seiten in den Suchergebnissen erscheinen.[17]

b) "Meta-Tag" zum Ausschluss der Indizierung

Wie gerade dargestellt, ist für ein „Verbot“ mittels /robots.txt ein Zugriff auf das Wurzelverzeichnis des Web-Angebots erforderlich. Jedoch wurden in den 1990er Jahren wegen der damals sehr hohen Preise für Domains Web-Angebote oftmals in Unterverzeichnissen abgelegt und gerade keine eigenen Domains genutzt. Daraus ergab sich das Problem, dass die Administratoren der einzelnen Angebote keinen Zugriff auf die Datei /robots.txt hatten. Hinzu kommt das vorstehend angesprochene Problem, dass Webseiten oder Inhalte trotz „Verbots“ in der /robots.txt weiterhin in den Suchergebnissen erscheinen können, wenn sie aufgrund anderer Hinweise als zur Suchanfrage passend eingestuft werden. Daher sieht eine Weiterentwicklung des *Robots Exclusion Protocol* aus dem Jahr 1996 vor, dass jede einzelne WWW-Seite Anweisungen an die Crawler enthalten kann. Realisiert wird dies durch Meta-Tags im sog. HTML-Kopf (*HEAD*), der hauptsächlich technische oder dokumentarische Informationen enthält, die üblicherweise nicht im Anzeigebereich des Browsers dargestellt werden.[18] Meta-Tags enthalten ergänzende Informationen zu einer Webseite, etwa eine Zusammenfassung oder Stichworte. Diese Funktion nutzt beispielsweise das Hamburger Abendblatt, um sein Archiv alter Texte aus Suchmaschinen herauszuhalten. Beispiel von http://suche.abendblatt.de:

```
<meta name="robots" content="noindex,follow,noodp"/>
```

Der Wert „noodp“ im Angebot des Hamburger Abendblatts weist die Suchmaschine an, nicht Titel und Beschreibung aus dem Open Directory Project zu verwenden.[19]

Sollen Regeln nur für bestimmte Crawler gelten, ist auch dies möglich, etwa zum Ausschluss des Google-News-Crawlers:

```
<meta name="googlebot-news" content="noindex" />
```

[16] https://support.google.com/webmasters/answer/93708.
[17] Vgl. https://support.google.com/webmasters/answer/93708.
[18] Vgl. zum Standard von Meta-Tags: http://www.robotstxt.org/meta.html.
[19] Zumindest die Suchmaschinen von Microsoft, Google und Yahoo! beachten diese Anweisung nach Kenntnis der DGRI auch.

Vorteil dieser Lösung ist, dass keine zentrale /robots.txt gepflegt werden muss und dennoch ganz spezifisch einzelne Seiten von der Indizierung ausgenommen werden können. Das entsprechende Meta-Tag kann für jede einzelne Seite unterschiedlich gesetzt werden.

Zudem entfernt jedenfalls Google derart gekennzeichnete Seiten nach eigener Darstellung vollständig aus seinen Suchergebnissen, im Gegensatz zur reinen Verhinderung des Abrufs der Dateien mittels /robots.txt.[20]

c) "X-Robots-Tag" im HTTP-Header zum Ausschluss der Indizierung

Das unter B. 3. b) beschriebene Meta-Tag ist für HTML-Seiten eine einfach umzusetzende Möglichkeit, Suchmaschinen eine Indizierung der jeweiligen Web-Seite zu untersagen. Das Konzept der Meta-Tags findet allerdings seine Grenzen, wenn es um andere Dateitypen wie etwa PDF-Dokumente oder Bilder geht. Jedoch gibt es die Möglichkeit, dem Crawler ein Indizierungsverbot nicht nur auf der Inhaltsebene im Meta-Tag in der HTML-Datei zu übermitteln, sondern auch auf der Protokollebene. Ein Eintrag auf der Protokollebene ist im Vergleich zur Verwendung von Meta-Tags (vgl. B. 3. b)) insofern von Vorteil, als er auf alle Dateitypen anwendbar ist und trotzdem zum selben Ergebnis führt:

Das WWW basiert auf dem HTTP(S)-Protokoll, das die Übertragung von Dateien (meist WWW-Seiten, zunehmend auch andere Daten) über ein Netzwerk (meist das Internet) regelt. Das HTTP-Protokoll sieht einen HTTP-Header mit Verwaltungsinformationen vor – vom Prinzip her ähnlich dem HTML-Header, der für den Nutzer ebenso unsichtbar bleibt –, der auch ein Header-Feld namens „X-Robots-Tag“ enthalten kann.

Diese Funktion nutzt beispielsweise die Seite http://www.spiegel.de/spiegel/print/d-14355425.html der Beklagten des Ausgangsverfahrens (diese Seite ist das erste Ergebnis, wenn man bei Google nach „Paul T.“ sucht). Dort enthält der HTTP-Header das folgende Feld:

```
X-Robots-Tag: index, follow, noarchive
```

Allerdings lautet die Anweisung beim Spiegel-Archiv ausdrücklich, die Seite in den Index aufzunehmen, nicht sie auszuschließen.[21]

Auch im HTTP-Header lassen sich die Anweisungen auf einzelne Crawler beschränken, etwa:

```
X-Robots-Tag: googlebot: noindex
```

Auf der vorstehend genannten WWW-Seite mit dem Text über den Beschwerdeführer wird auch eine PDF-Datei mit einem Scan der Original-Spiegel-Veröffentlichung verlinkt. Ruft man diese[22] auf, lautet das HTTP-Header-Feld:

```
X-Robots-Tag: noindex, nofollow, noarchive
```

Der Eintrag „noarchive“ bedeutet eine Anweisung an die Suchmaschinen, in den Ergebnislisten keine Kopie der Seite anzubieten.

Die Beklagte des Ausgangsverfahrens hat sich also bewusst entschieden, die HTML-Dateien ihres Archivs von Suchmaschinen indizieren zu lassen – „X-Robots-Tag: index...“ – , die PDF-Dateien des Archivs dagegen nicht – „X-Robots-Tag: noindex...“. Hintergrund dürfte sein, dass auf den HTML-

[20] https://support.google.com/webmasters/answer/93708.
[21] Im HTTP-Header: „index“ statt „noindex“.
[22] http://wissen.spiegel.de/wissen/image/show.html?did=14355425&aref=image036/2006/06/20/cq-sp198204701150122.pdf&thumb=false.

Seiten das sonstige Angebot von Spiegel Online verlinkt ist und Werbung geschaltet wird, während beim Aufruf einer PDF-Datei der Besucher „verloren“ ist, weil ihm ausschließlich der Content (die PDF-Datei) angezeigt wird.

d) Kombination mehrerer Maßnahmen

Im Folgenden werden die Kombinationen der Maßnahmen nach B. 3. a) bis c) hinsichtlich ihres Mehrwerts, sich vor Suchmaschinen zu „verstecken“, untersucht:

Eine Kombination eines Verbots mittels /robots.txt mit einem Verbot mittels Meta-Tag führt dazu, dass der Crawler die WWW-Seite wegen des /robots.txt Eintrags nicht mehr aufruft und so auch das Meta-Tag nicht mehr findet. Ein Verbot des Abrufs mittels /robots.txt geht somit technisch dem Verbot der Indizierung mittels Meta-Tag vor und die Kombination weist in diesem Fall keinen Mehrwert auf. Eine Kombination wäre nur dann förderlich, wenn ein Seitenbetreiber sicherstellen möchte, dass auch Suchmaschinen, die /robots.txt ignorieren, nicht aber Meta-Tags, die Indizierungsanweisungen beachten; dies jedoch um den Preis, dass gerade die seriösen Suchmaschinen, die sich an die /robots.txt-Anweisungen halten, die Seiten über Verlinkungen von Dritt-Seiten dennoch indizieren (vgl. oben B. 1. a. E.).

Auch bei einer Kombination von X-Robots-Tag im HTTP-Header mit /robots.txt gilt – wie beim HTML-Meta-Tag –, dass die WWW-Seite trotzdem (eingeschränkt) gefunden werden kann, denn der Crawler ruft die WWW-Seite nicht mehr direkt auf, so dass er den HTTP-Header-Eintrag nicht mehr finden kann.[23] Auch ein Verbot des Abrufs mittels /robots.txt geht somit etwaigen Anweisungen zur Indizierung mittels HTTP-Header-Feld technisch vor. Ein Mehrwert entstünde – wie oben schon dargestellt – nur, wenn eine Suchmaschine die /robots.txt nicht beachten würde.

Eine Kombination von X-Robots-Tag und Meta-Tag ist hinsichtlich der Nicht-Auffindbarkeit bei Suchmaschinen in der Regel nicht sinnvoll. Wie oben unter B. 3. c) aufgezeigt, liegt der Vorteil der Verwendung von X-Robots-Tags darin, dass sie aufgrund ihrer Funktionsweise im Gegensatz zu Meta-Tags alle Dateitypen erfassen, trotzdem aber dieselben Ergebnisse erbringen. Lediglich bei Suchmaschinen, die (nur) einen dieser Befehle missachten, kann eine Kombination der Maßnahmen der bessere Weg sein, um sich vor Suchmaschinen zu schützen. Zusammenfassend lässt sich zu den Kombinationen der Maßnahmen nach B. 3. a) bis c) sagen, dass diese zwar möglich sind, jedoch im Hinblick auf ihren nur geringen Mehrwert nicht zwingend von Nöten sind.

4. Nichtbereitstellen von Inhalten

Neben der Möglichkeit, Anweisungen an die Suchmaschinen mittels der oben dargestellten Befehle zu erteilen, besteht noch die weitaus sicherere und zeitlich schon früher greifende Methode, die Inhalte der Suchmaschine von vornherein überhaupt nicht zur Verfügung zu stellen. Mittels dieser Methoden, die entweder an den User Agent String, die IP-Adresse der Suchmaschinen oder an Passwörter anknüpfen, wird veranlasst, dass der Crawler der Suchmaschine schon gar nicht auf die Inhalte zugreifen kann und diese demzufolge nicht im Index erscheinen. In technischer Hinsicht überschneidet sich diese Variante mit den Varianten des direkten Sperrens einzelner Webinhalte, die unter C. angesprochen werden – an dieser Stelle soll lediglich auf technische Lösungen eingegangen werden, die Webinhalte spezifisch gegen den Zugriff von Suchmaschinen schützen.

[23] https://developers.google.com/webmasters/control-crawl-index/docs/robots_meta_tag?hl=de.

a) Sperrung aufgrund des *User Agent String*

Beim Abruf einer WWW-Seite sendet der abrufende Computer üblicherweise den sogenannten *User Agent String* an die aufzurufende Webseite mit.[24] Der User Agent String ist eine Angabe, welches Programm den Abruf vornimmt. Anhand des User Agent Strings werden in der Praxis so genannte Browser-Weichen realisiert, die dafür sorgen, dass jedem Browser (z.B. Firefox, Internet Explorer, Chrome, Opera) eine optimal angepasste WWW-Seite präsentiert wird. Eine andere Anwendung besteht beispielsweise darin, dass dem Nutzer ein Warnhinweis präsentiert wird, wenn er einen veralteten und damit für Angriffe anfälligen Browser verwendet.

Der Crawler von Googles Websuche beispielsweise verwendet die User Agent Strings

```
Mozilla/5.0 (compatible; Googlebot/2.1; +http://www.google.com/
bot.html)
```

und

```
Googlebot/2.1 (+http://www.google.com/bot.html)
```

während der Crawler für „Google News“ sich als

```
Googlebot-News
```

identifiziert.[25] Zugriffe durch den Googlebot lassen sich also am sichersten mithilfe des User Agent String „Googlebot“ identifizieren.[26] Ein Beispiel für den User Agent String für die Anfrage über einen Firefox-Browser wäre

```
Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:26.0)
Gecko/20100101 Firefox/26.0
```

ein Beispiel für Opera

```
Opera/9.80 (X11; Linux i686) Presto/2.12.388 Version/12.16
```

Im Internet gibt es umfassende Listen, welcher Crawler welchen User Agent String verwendet.[27] Fast alle Crawler benutzen das Schlagwort „Bot“ in ihrem User Agent String, aber auch „Spider“ hat eine gewisse Verbreitung. Beide Schlagworte kommen, soweit für uns ersichtlich, in den User Agent Strings normaler Browser nicht vor, so dass bereits eine einfach Abfrage[28] im Content Management System des Anbieters die gängigen Suchmaschinen identifizieren sollte.[29]

Es wäre also beispielsweise technisch möglich, abhängig vom anfragenden User Agent String, entweder (i) den regulären Text der WWW-Seite oder (ii) eine Fehlermeldung oder (iii) eine leere Seite auszuliefern. Diese Technik wird auch von DGRI-Mitgliedern verwendet, um bestimmte Inhalte Suchmaschinen vorzuenthalten, ohne normale Nutzer zu beeinträchtigen.

---

[24] Vgl. RFC 2616, Abschn. 14.43.
[25] https://support.google.com/webmasters/answer/1061943?hl=de&ref_topic=2370570.
[26] https://support.google.com/webmasters/answer/80553?hl=de.
[27] Vgl. hierzu nur: http://www.useragentstring.com/pages/Crawlerlist/.
[28] Abfrage würde vereinfacht dargestellt in etwa lauten: „Enthält der User Agent String die Zeichenfolgen ‚Bot‘ oder ‚Spider‘?“.
[29] Um auch den Yahoo!-Crawler mit einzubeziehen, wäre ergänzend noch nach den Zeichenfolgen „Yahoo“ oder „Slurp“ zu suchen.

Hinweis: Es ist möglich, durch entsprechende Einstellungen bzw. Zusatzsoftware den User Agent String zu verändern und sich beispielsweise als „Googlebot“ auszugeben. Diese Nutzer würden dann ebenso wie eine Suchmaschine behandelt. Umgekehrt lässt sich natürlich nicht ausschließen, dass sich eine Suchmaschine als normaler Browser tarnt,[30] eine solche Suchmaschine dürfte aber praktisch keine Bedeutung am Markt haben und damit auch keine Bedeutung für das streitgegenständliche Problem des automatisierten Zugriffs auf alte Presseberichte.

b) Sperrung anhand der IP-Adresse des Anfragenden

Alternativ oder ergänzend neben der Anknüpfung an den User Agent String ist auch eine Filterung anhand der IP-Adresse denkbar: Der Google-Crawler beispielsweise verwendet oftmals IP-Adressen, die aus dem Netz „66.249.x.y“ stammen und deren Name auf „googlebot.com“ endet. Die Crawler der Suchmaschine Baidu beispielsweise verwenden Namen, die auf „crawl.baidu.com“ oder „crawl.baidu.jp“ enden.[31] Entsprechend dürften sich – mit für uns schwer einzuschätzendem Aufwand – für die meisten Suchmaschinen jedenfalls große Teile der aktuellen IP-Adressbereiche herausfinden lassen, um diese zu blocken. Google stellt jedoch keine öffentliche Liste der IP-Adressen zur Verfügung. Begründet wird das Vorenthalten der IP-Adressen damit, dass sich diese leicht ändern können.[32] Somit kann diese Maßnahme alleine keinen geeigneten Schutz gegen Suchmaschinen erbringen, sondern müsste noch mit anderen Maßnahmen kombiniert werden, um ausreichenden Schutz zu gewährleisten.

c) HTTP-Authentifizierung zur Zugriffsverhinderung

Die HTTP-Standardauthentifizierung[33] ist die einfachste Technik zur Durchsetzung von Zugriffskontrollen auf Web-Ressourcen, also ein Verfahren, mit dem sich der Nutzer eines Webbrowsers gegenüber dem Webserver bzw. einer Webanwendung authentifizieren kann, um anschließend für weitere Zugriffe autorisiert zu sein. Durch einen solchen Schutzmechanismus kann es Crawlern unmöglich gemacht werden, auf Inhalte zuzugreifen, die sich auf dem Web-Server in einem solchen passwortgeschützten Verzeichnis befinden bzw. deren Dateiname einem festgelegten Muster entspricht. Dies ist die effektivste Maßnahme, um eine Indizierung von Inhalten zu verhindern[34]. Ein Crawler würde an der Autorisierungsabfrage beim Aufruf eines der Verzeichnisinhalte scheitern. Allerdings müsste auch jeder legitime Nutzer sich mittels Benutzernamen und Passwort authentifizieren, was für frei zugängliche Inhalte kaum in Betracht kommen dürfte.

5. Löschungsantrag bei Suchmaschinenbetreibern

Google, aber auch andere Suchmaschinenbetreiber bieten die Möglichkeit, für eigene oder fremde Webseiten einen Antrag zu stellen, dass diese nicht mehr in der Trefferliste geführt werden, also aus dem Trefferpool der Suchmaschinen gelöscht werden.[35] Allerdings löscht Google nur ganz wenige Arten von Inhalten, etwa Kreditkartennummern oder Bilder von Kindesmissbrauch,[36] oder wenn be-

---

[30] Im Internet werden verschiedene sogenannte „Peer-to-peer Search Engines“ bereitgestellt (z.B.: http://yacy.net/en/), die wahlweise als Crawler oder über Proxy Server arbeiten, also Rechner von „Mitgliedern“ für Abfragen nutzen, die dieses Programm bei sich installiert haben. Soweit die Abfrage von einem Proxy Server ausgeht, ist denkbar, dass das Programm den User Agent String des Proxy Servers nutzt. Wegen der zu vernachlässigenden wirtschaftlichen Bedeutung dieser Search Engines haben wir bisher von einer näheren Prüfung abgesehen.

[31] Vgl. http://help.baidu.com/question?prod_en=master&class=498&id=1000973.

[32] Vgl. https://support.google.com/webmasters/answer/80553?hl=de.

[33] Vgl. RFC 2617.

[34] Vgl. https://support.google.com/webmasters/answer/93708.

[35] Vgl. für Google: https://support.google.com/websearch/troubleshooter/3111061.

[36] https://support.google.com/websearch/answer/2744324.

reits eine gerichtliche Entscheidung gegen den Anbieter der Website ergangen ist.[37] Die im vorliegenden Fall relevanten Inhalte gehören nicht dazu, jedenfalls nicht, solange gegen den Anbieter kein gerichtliches Verbot ergangen ist.
Sind die Inhalte bereits von der originalen Website entfernt worden, erscheinen sie aber noch in den Google-Suchergebnissen,[38], kann über die „Google Webmaster-Tools“ eine vorzeitige Löschung beantragt werden.[39] Diese Maßnahme erscheint daher nur als Zusatzmaßnahme begleitend zu den Maßnahmen nach B. 3. a) bis c) sinnvoll. Zudem sei darauf hingewiesen, dass im Gegensatz zu den Hauptmaßnahmen nach B. 3. a) bis c) diese Begleitmaßnahme auf Antrag beim Suchmaschinenbetreiber erfolgt und daher auch von diesem durchgeführt werden muss, also ein zeitliches Restrisiko hinsichtlich der Erledigung durch den Suchmaschinenbetreiber besteht.

6. Zusammenfassung zu den Möglichkeiten von Internet-Portalen, auf die von Suchmaschinen aufgefundenen und angezeigten Ergebnisse Einfluss zu nehmen

Es ist Betreibern von WWW-Seiten grundsätzlich ohne Weiteres möglich, bestimmte Inhalte von der Indizierung durch die marktrelevanten Suchmaschinen auszuschließen, ohne die Anzeige der Web-Inhalte für andere Besucher in irgendeiner Form zu beeinträchtigen. Dazu stehen verschiedene Möglichkeiten zur Verfügung, die in der Praxis auch bereits umfassend genutzt werden. Zu diesen Nutzern gehört auch die Beklagte des Ausgangsverfahrens – auch wenn diese die aufgezeigten Möglichkeiten teilweise gerade dazu nutzt, um ein Erscheinen in den Suchmaschinen zu erreichen und nicht das vom Beschwerdeführer gewünschte Nichterscheinen der Berichte über ihn.

Bezüglich des damit verbundenen Aufwands sei auf die folgenden Ausführungen verwiesen. Insbesondere ist zu beachten, dass Anwender aus Gründen des Aufwands dazu motiviert sein könnten, die vorgestellten Maßnahmen unabhängig vom Inhalt der WWW-Seiten anzuwenden, um sich den Aufwand einer Prüfung im Einzelfall zu ersparen.

[37] Formular zur Einreichung unter https://support.google.com/legal/contact/lr_courtorder?product=websearch.
[38] Aufgrund der Größe der Datenmengen im Internet benötigt der Crawler einige Zeit, bis er jede Webseite wieder „besucht“ hat. Allerdings ist nicht öffentlich bekannt, wie lange der Crawler benötigt, bis er die Seiten der Beklagten wieder „besucht“; erfahrungsgemäß ist die Besuchsfrequenz bei regelmäßig aktualisierten Inhalten wie Nachrichten-Websites erheblich höher als bei statischen Angeboten. Allerdings werden in PDF-Form vorliegende Inhalte von den Suchmaschinen nur dann erneut indiziert, wenn der Name der Datei geändert wurde, vgl. C. 2. e).
[39] https://support.google.com/legal/troubleshooter/1114905?rd=1#ts=1115655,1282866.

C. Möglichkeiten und Aufwand, um den direkten Zugriff auf personenbezogene, in einem online zugänglichen Archiv gespeicherte Daten zu verhindern oder einzuschränken

1. Einflussnahme auf Suchmaschinen

Sämtliche unter B. vorgestellten Maßnahmen schränken den Zugriff auf personenbezogene, in einem Online-Archiv gespeicherte Daten massiv ein, soweit der Interessent nicht direkt auf die Inhalte zugreift, sondern eine allgemeine Suche mittels einer der üblichen Suchmaschinen durchführt. Mit Ausnahme der Zugriffskontrolle via der Datei /robots.txt und einer Anonymisierung, die durch Zusatzinformationen aus anderen Quellen unterlaufen werden können (vgl. B. 1. am Ende), dürften in der Praxis bereits nach wenigen Tagen bis Wochen die für Suchmaschinen gesperrten Inhalte nicht mehr mittels Suchmaschinen aufzufinden sein.

Keinerlei Einfluss haben die unter B. 3. vorgestellten Maßnahmen (anders als die unter B. 2. Angesprochene Möglichkeit die Beiträge selbst inhaltlich zu verändern und so das veröffentlichte Wissen faktisch allgemein unerreichbar zu machen) allerdings auf die grundsätzliche Zugänglichkeit der Daten, wenn der Interessent die Webseite des Anbieters unmittelbar aufsucht und auf die streitigen Web-Inhalte entweder direkt oder über den Weg einer Suchfunktion auf der Web-Seite des Online-Anbieters[40] oder auch einer kostenpflichtigen Datenbank wie beispielsweise Genios zugreift. So lässt sich z.B. auf der Archiv-Seite des Hamburger Abendblattes der Fall des Beschwerdeführers problemlos anhand der Original-Texte nachvollziehen. In den Suchergebnissen erscheinen diese jedoch nicht, da der Verlag das Meta-Tag „noindex“ gesetzt hat.

2. Aufwand der Einflussnahme auf Suchmaschinen

Die Kosten der technischen Umsetzung der unter B. beschriebenen Maßnahmen beschränken sich auf die Kosten der einmaligen Einrichtung der Maßnahmen. Die zusätzlichen Kosten für Datenspeicherung und -verkehr dürften nicht messbar sein. Bereits aus Eigeninteresse nutzen allerdings viele Medien die aufgezeigten Maßnahmen bereits, weil dadurch eine Steuerung der Darstellung in den Suchmaschinen möglich ist. Zwar sollten somit die technischen Grundvoraussetzungen in aller Regel schon vorhanden sein und kostentechnisch nicht allzu stark ins Gewicht fallen, jedoch kommen noch in jedem Fall die Kosten des internen Entscheidungsprozesses hinzu.

a) Zugriffsbeschränkung des gesamten Archivs

Geht es um eine pauschale Umstellung des gesamten Archivs - müsste also beispielsweise bei der Beklagten des Ausgangsverfahrens der Administrator die Konfigurationsdatei für den HTTP-Header-Eintrag dergestalt ändern, dass es statt „index“ künftig „noindex“ heißt -, dürfte der Umsetzungsaufwand abhängig von den Zugangshürden zum Server zwischen wenigen Sekunden und wenigen Minuten liegen. Ähnliches gilt für die anderen unter B. 3. genannten Maßnahmen, soweit sie den Zugriff auf die Webseite insgesamt beschränken sollen. Eine Implementierung der unter B. 4. a) und b) genannten „Weichen“ erfordert Programmieraufwand, dessen Umfang wesentlich von der verwendeten Software abhängen dürfte. Eine Benutzer-Authentifizierung (B. 4. c)) scheidet für an die Öffentlichkeit gerichtete WWW-Seiten faktisch aus; will man sie dennoch verwenden, verursacht sie entweder kaum nennenswerten Aufwand (wenn für alle Nutzer ein Benutzername nebst Passwort bereitgestellt wird; dies erfordert einen kurzen Eintrag in der Konfiguration) oder einen erheblichen Aufwand (wenn sich jeder Nutzer registrieren und eine individuelle Kombination aus Benutzernamen und Passwort erhalten müsste).

---

[40] Bei dieser Suchfunktion muss es sich um eine eigene, integrierte Funktion des Betreibers der Webseite handeln und nicht um eine Suchfunktion, die mit Hilfe der gängigen Suchmaschinen arbeitet.

b) Herausnahme nur einzelner Dateien aus der Indizierung

Sollen dagegen nur einzelne Dateien aus der Indizierung durch die Suchmaschinen herausgenommen werden, kann der Aufwand erheblich höher sein. Wie hoch dieser Aufwand wäre, lässt sich nur in Kenntnis der konkreten technischen Gegebenheiten bei dem jeweiligen Online-Anbieter abschätzen:

(1) Sieht die vom Online-Anbieter verwendete Software beispielsweise bereits die Möglichkeit vor, für einzelne Internetseiten gesonderte HTML-Meta-Tags (vgl. oben B. 3. b)) in den Seitenquelltext einzubauen, ist der mit der Anpassung der betroffenen Seite im Einzelfall verbundene Aufwand sehr gering und dürfte einige Sekunden bis wenige Minuten pro Text nicht überschreiten. Folglich drohen hier „nur“ Summierungs-Effekte (durch die Zahl individueller Löschungs-/Sperrungsanträge).

Muss dagegen zunächst die verwendete Software umprogrammiert werden, können die Kosten je nach Komplexität der Software erheblich sein.

(2) Ein HTTP-Header-Eintrag (vgl. oben B. 3. c)) dagegen lässt sich etwa bei den weit verbreiteten „Apache“-Webservern durch Bearbeitung einer Textdatei mit sehr geringem Aufwand erstellen. Hier stellt sich allerdings die Frage, ob sich bei einer sehr großen Zahl von Einträgen für einzelne Webseiten bzw. Dateien Auswirkungen auf die Geschwindigkeit des Servers ergeben könnten, weil diese Datei bei jeder Anfrage ausgewertet werden muss. Nach – nicht zwingend repräsentativen – Berichten von Administratoren seien auch bei sehr großen Konfigurationsdateien keine Auffälligkeiten festzustellen. Nachdem die Anfrage-Belastung der Server großer Anbieter wie der Beklagten des Ausgangsverfahrens eine ganz erhebliche sein dürfte, können wir insoweit jedoch keine belastbare Aussage treffen, umso mehr, als es in diesem Bereich diverse Optimierungsmöglichkeiten gibt, die je nach dem verwendeten System zu berücksichtigen sein könnten. Klarheit dürfte hier wohl nur ein Test im Produktivbetrieb eines großen Anbieters schaffen.

Wenig problematisch dürften dagegen Fälle sein, in denen die URL der von der Indizierung auszunehmenden Dateien einem bestimmten Muster folgt – etwa alle PDF-Dateien oder alle Dateien aus dem Verzeichnis „/2014/“ –, denn durch Umsetzen dieses Musters in einen einzigen Befehl würden sich alle betroffenen Dateien auf einmal beeinflussen lassen.

c) Entfernung bzw. Anonymisierung einzelner Beiträge aus einem Redaktionssystem

Einen bestimmten, konkret etwa durch eine Abmahnung bezeichneten Beitrag vollständig aus einer Datenbank zu löschen, dürfte nur einen geringen Aufwand von einigen Sekunden bis maximal wenigen Minuten verursachen, je nach Ausgestaltung im Einzelfall. Im schlimmsten Fall muss der Beitrag im System aufgerufen, manuell gelöscht und anschließend ohne Inhalt erneut gespeichert werden.

Typischerweise dürfte die von Presse-Archiven verwendete Software zudem ein Feld für die Sperrung einzelner Beiträge vorsehen, weil eine Sperrung aufgrund von gerichtlichen Verboten oder Unterlassungserklärungen ohnehin ein Standardfall ist. Der Beitrag wäre dann nicht gelöscht – und unter Umständen auch redaktionsintern weiterhin verfügbar, ggf. mit einem Hinweis zu Tatsache und Grund der Sperrung –, aber der Öffentlichkeit nicht mehr zugänglich. Eine solche Sperrung dürfte je nach Ausgestaltung im Einzelfall nur Sekunden in Anspruch nehmen. Vorteil dieser Methode wäre, dass diese Beiträge weiterhin zu internen Recherchen zur Verfügung stünden und somit die Eingriffsintensität in pressespezifische Tätigkeiten verringert würde.

Soll nicht der gesamte Beitrag gelöscht oder gesperrt werden, sondern, etwa im Hinblick auch auf das Grundrecht der Informationsfreiheit, nur eine Anonymisierung des Betroffenen erfolgen, gilt:

(1) Eine solche Bearbeitung könnte zunächst per Hand erfolgen, indem ein Mitarbeiter den betroffenen Text durchliest und die betreffenden Stellen entfernt. Soll nur ein konkreter Begriff entfernt werden, beträgt der Aufwand – abhängig von der Länge des Textes – einige Minuten pro Beitrag. Jedenfalls wenn Rechtsfolgen an das Nichtentfernen von bestimmten Angaben geknüpft werden sollen, sollte wegen der zu erwartenden Fehlerquoten der Zeitaufwand für eine zusätzliche manuelle Überprüfung zur Kontrolle des Löscherfolgs über die Suchfunktion des Online-Angebots und ggf. erforderliches Nacharbeiten hinzuzurechnen sein.

(2) Sollen nur bestimmte Zeichenfolgen, etwa der Nachname einer Person, aus einem Text entfernt werden, kann dies zudem mit der üblichen Funktion „Suchen und ersetzen“ erfolgen. Falls das Redaktionssystem hierfür keine Funktion vorsieht, kann der betroffene Text typischerweise jedenfalls aus dem Redaktionssystem in eine Textverarbeitung kopiert, dort der Ersetzungsvorgang vorgenommen und der bearbeitete Text sodann wieder ins Redaktionssystem kopiert werden. Abhängig davon, ob der Beitrag auch Formatierungsbefehle oder sonstige Zusatzinformationen enthält und in welcher Form diese vorliegen, kann das Verfahren mittels Textverarbeitungsprogramm allerdings ausscheiden, wenn diese Zusatzinformationen nicht fehlerfrei übernommen werden können. In diesem Fall müsste das verwendete Redaktionssystem selbst eine Funktion „Suchen und ersetzen“ vorhalten, oder es müsste die händische Bearbeitung gemäß C. 2. c) (1) gewählt werden.
Problematisch bleibt, dass bei solchen automatischen Funktionen nur die exakt definierte Zeichenfolge ersetzt wird, dies jedoch stets. Dieser Umstand kann im Einzelfall zu überzähligen oder fehlenden Ersetzungen führen (siehe dazu die Beispiele sogleich unter C. 2. d)). Solche überzähligen Ersetzungen lassen sich indes, unter Inkaufnahme eines zeitlichen Mehraufwandes abhängig von der Zahl der einzelnen Ersetzungen, vermeiden, wenn das Ersetzen nicht automatisch für den gesamten Text vorgenommen wird, sondern der Bearbeiter nach inhaltlicher Bewertung im Einzelfall jeden einzelnen Ersetzungsvorschlag individuell freigibt (vgl. dazu auch im Folgenden unter C. 2. c) (3)).

(3) Soll auch eine sonstige Identifizierbarkeit einer Person unterbunden werden – wie dies von der presserechtlichen Rechtsprechung regelmäßig verlangt wird –, wird regelmäßig eine umfassende redaktionelle Bearbeitung des Textes erforderlich sein. Dies erfordert regelmäßig umfassende Kenntnis des konkreten Falls, um die problematischen, eine Identifizierung ermöglichenden Aspekte der Berichterstattung erkennen und verändern zu können. Ein solches umfassendes inhaltliches Redigieren eines längeren Beitrags kann (für jeden einzelnen Beitrag) mehrere Stunden dauern. Muss der Bearbeiter sich zudem erst in den Sachverhalt einarbeiten, was bei älteren Fällen vielfach vorkommen kann, erhöht sich der Zeitbedarf nochmals erheblich.

d) Aufwand für die beitragsübergreifende Veränderung oder Sperrung aller Beiträge mit bestimmten Inhalten

Soweit – wie üblich – über konkret (namentlich mit Fundstelle/URL) benannte Veröffentlichungen hinaus gar sämtliche identifizierende Berichterstattung über bestimmte Umstände unterbunden werden soll, erfordert dies eine umfassende Veränderung oder Sperrung sämtlicher Beiträge, die dem Verbot unterfallende Inhalte enthalten, ob diese nun im jeweiligen Rechtsstreit konkret benannt wurden oder nicht. Soweit der Betreiber auf ein derartiges Verbot nicht mit einer Sperrung oder Löschung seines gesamten (Archiv-) Angebots reagieren will (was, wie unter C. 2. a) ausgeführt, mit verhältnismäßig geringem Aufwand verbunden wäre, aber schwerwiegende Auswirkungen auf die Grundrechte des Art. 5 GG hätte), bedeutet ein derartiges Verbot erheblichen zusätzlichen Aufwand. Denn in diesem Fall müsste der Seitenbetreiber über eine Recherche zunächst alle Beiträge identifizieren, die vom Verbot betroffene Inhalte aufweisen (könnten). Dies wird, auch bei Inanspruchnahme technischer Hilfsmittel bei der Identifizierung möglicher relevanter Beiträge, nicht ohne eine redaktionelle Überprüfung, durchgeführt werden können - die wiederum Kenntnis des konkreten Falles voraussetzt.

Dies gilt unabhängig davon, ob der Betreiber der Webseite die Entfernung oder Änderung solcher Inhalte (vgl. oben B. 2.) oder Anweisungen an Suchmaschinen einrichten will, die den Zugriff auf konkrete Beiträge verhindern oder beschränken (vgl. oben B. 3. und B. 4.).

Falls das verwendete Redaktionssystem eine Funktion zum beitragsübergreifenden Suchen und Ersetzen bestimmter Zeichenketten aufweist, kommt immerhin dessen Nutzung zur Identifizierung möglicherweise vom Verbot betroffener Beiträge und zu deren Änderung in Betracht.

Dabei ist jedoch zu berücksichtigen, dass – jedenfalls ältere – Presseveröffentlichungen in allen der DGRI bekannten Fällen mittels Einscannen und Zeichenerkennung digitalisiert wurden. Dieses Verfahren ist zum einen grundsätzlich fehleranfällig, weil die Fehlerquote beim Übertragen des Ursprungstextes in ein Textformat (sog. OCR-Erkennung) von einer Vielzahl von Faktoren abhängt, etwa von Fehlern, Beschädigungen oder Verschmutzungen des Originaldokuments, der Qualität des Scans oder auch der zu suchenden Zeichenkette. Insbesondere (Nach-)Namen weisen dabei typischerweise eine erheblich höhere Fehlerquote auf als Worte des normalen deutschen Wortschatzes, denn Zeichenerkennungssoftware versucht im Fall nicht eindeutiger Ergebnisse durch Abgleich mit Wörterbüchern das nächstliegende sinnvolle Wort zu finden. Gibt es – wie bei Nachnamen üblich – keinen Wörterbuch-Eintrag, muss sich die Software auf ihre Erkennungsleistung verlassen. Da bereits Menschen – insbesondere bei Ligaturen oder eng zusammenstehenden, aber ähnlich aussehenden Zeichenfolgen wie 'rn' und 'm' – Schwierigkeiten haben, bestimmte Schriften korrekt zu lesen, sind Fehler bei Nachnamen vorprogrammiert. Hinzu kommt, dass mangels Wörterbuch-Unterstützung auch Trennzeichen nach dem bisherigen Stand der Technik offenbar nicht automatisch entfernt werden können: Der im „Spiegel“-Archiv zu findende „Jürgen Pe-t.“ könnte schließlich tatsächlich so heißen. Bereits eine automatische Identifizierung ist damit nicht sicher möglich; selbst bei der Nutzung „unscharfer“ Suchalgorithmen erscheint ein Auffinden sämtlicher Fundstellen nicht sichergestellt.

Im Fall der Anonymisierung des Namens des Betroffenen in identifizierten betroffenen Beiträgen könnte auch eine ggf. vorhandene Funktion „Ersetzen“ genutzt werden. Eine solche Änderung würde – bei Vorhandensein einer das gesamte Redaktionssystem umfassenden Ersetzungsfunktion – daher typischerweise maximal wenige Minuten in Anspruch nehmen. Allerdings dürfte hier eine gesonderte manuelle Überprüfung erforderlich sein, ob die vorgenommenen Änderungen inhaltlich zutreffend sind. So ist etwa denkbar, dass – um ein konkretes Beispiel zu wählen – das Medium an ganz anderer Stelle nicht nur über den Beschwerdeführer, sondern auch über einen „Hans T.“ berichtet hat, also Nachnamens-Identität zwischen den betroffenen Personen besteht. Würde man nun pauschal in allen Beiträgen „T.“ entfernen bzw. ersetzen, wären auch die – möglicher Weise zulässigen – Berichte über „Hans T.“ betroffen. Würde man dagegen nur nach dem Gesamtbegriff „Paul T.“ suchen, würden weder "Herr T." noch "T." in Alleinstellung noch „Paul-Friedrich T.“ gefunden.

Nicht vergessen werden darf auch die Problematik identischer Namen als weitere Erschwerung im Vergleich zur bloßen, oben schon beispielhaft dargestellten Nachnamensidentität unterschiedlicher Personen, die in jedem Fall eine manuelle Überprüfung der betroffenen Texte erfordert.

Die vorgenannten Probleme ergeben sich entsprechend, wenn es nicht um eine automatische Veränderung, sondern um eine automatisierte Löschung oder Sperrung konkreter Beiträge geht.

Im Ergebnis scheidet eine automatische Ersetzung, Sperrung oder Löschung über das gesamte Archiv praktisch aus, da zum einen dieses Vorgehen aufgrund der oben dargestellten Gründe zu fehleranfällig sowie in Spezialfällen zu weitgreifend sein kann und zum anderen der Aufwand dafür größer sein dürfte als die betroffenen Beiträge über die Suchfunktion zu identifizieren und manuell die relevanten Beiträge zu bearbeiten.

e) Aufwand für die Änderung von gescannten PDF-Dokumenten

Soweit es nicht um die Änderung von Texten in einer Datenbank geht, sondern um Änderungen an eingescannten PDF-Versionen alter Veröffentlichungen, kommt zu den vorgenannten Schwierigkeiten ein technisches Problem bei der Anonymisierung hinzu:

Zunächst wird ein Auffinden der betroffenen Texte überhaupt nur möglich sein, wenn die Suche auf Basis mittels OCR in Textform umgewandelter Artikel erfolgt, weil gescannte, nur mit einer bildlichen Wiedergabe der Ursprungsveröffentlichung versehene PDF-Dokumente nicht (maschinell) durchsuchbar sind.

Sodann müsste eine solche Änderung/Teillöschung in folgenden Schritten erfolgen: (i) Aus dem PDF-Dokument wird das Bild der Seite extrahiert, in der sich die zu anonymisierende Textstelle befindet; (ii) in diesem Bild werden mit einem Bildbearbeitungsprogramm die zu entfernenden Inhalte gelöscht/ geschwärzt und sodann ein neues PDF-Dokument erstellt, das wiederum in das Archiv eingestellt werden kann; (iii) dies muss unter einem neuen PDF-Dateinamen erfolgen, da Suchmaschinen davon ausgehen, dass PDF-Dokumente nicht geändert werden, sie also ein geändertes PDF-Dokument unter altem Dateinamen nicht erneut mittels OCR auslesen würden.
Das vorbeschriebene Vorgehen erscheint aus Sicht der DGRI die einzige technisch sichere Vorgehensweise; insbesondere die häufig zur Zeitersparnis verwendete Variante, im PDF-Dokument direkt mit Hilfe einer PDF-Bearbeitungssoftware einen "schwarzen Balken" zur Anonymisierung zu platzieren, ist keine sichere Variante, da sich diese Balken mit Hilfe ebensolcher Software auch wieder von jedem Dritten entfernen lassen.

Es ist zwar denkbar, für dieses Verfahren eigens Software zu programmieren, die diese Aufgabe weitgehend automatisiert; auch in diesem Fall dürften aber zumindest einige Minuten pro Beitrag Änderungsarbeit anfallen. Solange keine spezielle Software programmiert ist, kann der Zusatzaufwand – je nach Software-Ausstattung des Mediums – erheblich höher liegen.

3. Kosten des Entscheidungsprozesses

Die Kosten der rein technischen Umsetzung einer Löschung, Sperrung oder inhaltlichen Änderung eines Beitrags oder des Zugriffs darauf können sich als ein nur untergeordneter Aspekt der wirtschaftlichen Belastung des Online-Mediums darstellen. Wie insbesondere unter C. 2. d) bereits angedeutet, kann der wesentliche Aufwand in der Entscheidung bestehen, welche konkrete Information zu löschen, ändern bzw. sperren ist, bzw. auf welche konkrete Information der Zugriff verhindert bzw. beschränkt werden soll. Soweit dieser Aufwand einzelne Texte betrifft, ist dies im Rahmen der technischen Umsetzung angesprochen worden, weil dieser Aufwand auch dann anfällt, wenn einem Medium eine bestimmte Änderung, Löschung oder Sperrung konkret – etwa aufgrund einer Abmahnung oder einer Unterlassungsverpflichtung – aufgegeben wird. Dabei sei insbesondere daran erinnert, dass der übliche, nicht auf konkret benannte Veröffentlichungen beschränkte, sondern nur auf ein Thema bezogene Verbotsausspruch im Fall von Archiven dem betroffenen Medium praktisch nicht erfüllbare Verpflichtungen auferlegt (vgl. C. 2. c) (3) und C. 2. d)).

Soweit Online-Archive darüber hinaus dazu verpflichtet werden sollen, selbständig ihre Altveröffentlichungen auf zwischenzeitlich nicht mehr zulässige Berichterstattung zu überprüfen,[41] müsste jeder einzelne Altbeitrag vor dem Online-Stellen redaktionell-juristisch überprüft werden. Da sich ein Personenbezug bekanntermaßen nicht nur aus einer vollen Namensnennung, sondern auch aus einer sonst identifizierenden Berichterstattung ergibt, müsste die den Beitrag überprüfende Person sowohl

[41] So jedenfalls die Auffassung des Berufungsgerichts, OLG Hamburg, Urteil v. 01.11.2011 – 7 U 49/11, für den Zeitpunkt des erstmaligen Online-Stellens der Beiträge, auch wenn das Berufungsgericht später (Urteil vom 29.11.2011 – 7 U 80/11) nur noch auf den Zugang einer Abmahnung abstellt.

Fachkenntnisse über den Gegenstand der Berichterstattung haben als auch die rechtliche Kompetenz zur Bewertung, ob die Veröffentlichung (noch) zulässig ist.

Ergibt die Bewertung, dass eine identifizierende Altberichterstattung zum Zeitpunkt der Einstellung ins Online-Archiv zulässig ist, sagt dies nach der angesprochenen Rechtsauffassung noch nichts über die Dauer der Rechtmäßigkeit der Weiter-Veröffentlichung aus. Ebenso wie alle neu produzierten Beiträge müsste der gesamte Altbestand regelmäßig erneut redaktionell-rechtlich überprüft werden, um zwischenzeitlich rechtswidrig gewordene Veröffentlichungen zu identifizieren und zu entfernen.

Eine generelle Verpflichtung, Altberichterstattung aus Print-Medien bei der Aufnahme in ein Online-Archiv oder gar laufend redaktionell-rechtlich zu überprüfen, käme aufgrund des technischen und vor allem zeitlichen Aufwands einem faktischen Online-Veröffentlichungsverbot gleich. Diese These soll anhand des folgenden Zahlenbeispiels untermauert werden:

Das Online-Archiv des „Hamburger Abendblatts“ verzeichnet für die Ausgabe vom 6. Februar 2014 insgesamt 420 Beiträge. Das Online-Archiv des „Spiegel“ beinhaltet für Heft 6/2014 insgesamt 112 Beiträge. Das Online-Archiv der „taz“ enthält für den 6. Februar 2014 189 Beiträge. Alleine für die Jahre 1986 bis 2013 summieren sich die Beiträge aus der gedruckten „taz“ in deren Online-Archiv auf insgesamt 1.467.030 Texte. In den Zahlen nicht berücksichtigt ist eine eventuelle Bildberichterstattung einschließlich der dortigen (identifizierenden) Bildunterschriften.

Nimmt man für jeden einzelnen dieser 1.467.030 taz-Printausgaben-Texte einen Überprüfungsaufwand von nur fünf Minuten an, würde die Überprüfung 122.252,5 Stunden dauern, d.h. etwa 3.500 Arbeitswochen.[42] Dies entspricht bei durchschnittlich 14,2 Krankheitstagen pro Jahr[43] und angenommenen durchschnittlichen 32 Urlaubstagen pro Jahr[44] bei etwa 250 Arbeitstagen pro Jahr abzgl. Krankheits- und Urlaubstagen etwa 86 Mannjahren.

Das im Ausgangsverfahren vom Berufungsgericht[45] vorgebrachte Argument, es würde eine Überprüfung der Texte aller Gerichtsreporter genügen, berücksichtigt bereits nicht, dass die Rechtsprechung (zu Recht) nicht nur Straftätern persönlichkeitsrechtliche Abwehransprüche gegen identifizierende Berichterstattung zugesteht. Es berücksichtigt auch nicht, dass nicht nur Gerichtsreporter über Straftaten berichten, sondern auch die meisten sonstigen Journalisten zumindest gelegentlich, z.B. im Wirtschaftsressort, wenn die Straftat Auswirkungen auf ein Unternehmen hat.

4. Zusammenfassung zu den Möglichkeiten und dem Aufwand, um den direkten Zugriff auf personenbezogene, in einem online zugänglichen Archiv gespeicherte Daten zu verhindern oder einzuschränken

Altberichterstattung pauschal von der Indizierung durch Suchmaschinen auszunehmen und dadurch das Risiko der Auffindbarkeit bei pauschalen Recherchen nach einer Person in relevantem Umfang zu verringern, verursacht keine nennenswerten Kosten (C. 2. a)).

Der Aufwand, einzelne Beiträge von der Indizierung durch Suchmaschinen auszunehmen, kann je nach den beim Medium vorliegenden technischen Gegebenheiten im Einzelfall – insbesondere dem

[42] Unter Annahme einer 36,5 Stunden-Woche nach § 7 Abs. 1 Manteltarifvertrag für Redakteure und Redakteurinnen an Tageszeitungen – gerechnet mit 35 Stunden, weil mindestens 1,5 Stunden für die Koordination abzuziehen wären.
[43] Gesundheitsreport 2013, Techniker Krankenkasse, S. 73 – abrufbar unter: http://www.tk.de/centaurus/servlet/contentblob/516416/Datei/83065/Gesundheitsreport-2013.pdf.
[44] Vgl. § 9 Abs. 2 Manteltarifvertrag für Redakteure und Redakteurinnen an Tageszeitungen: 30 bis 34 Tage.
[45] OLG Hamburg, Urteil v. 01.11.2011 - 7 U 49/11, S. 7 des Urteils.

Vorhandensein entsprechender Mechanismen in der verwendeten Software – zwischen nicht nennenswert und erheblich liegen (C. 2. b)).

Der Aufwand, einzelne Beiträge generell für die Anzeige im Online-Archiv zu sperren, dürfte bei der üblicherweise eingesetzten Software gering sein (C. 2. c)).

Der Aufwand, einzelne Beiträge zu anonymisieren, kann je nach Art der Beanstandung (Namensnennung oder sonstige Identifizierbarkeit) und ggf. Funktionen der verwendeten Software zwischen einigen Minuten und Stunden liegen (C. 2. c) und C. 2. e)).

Zu beachten sind für alle Maßnahmen, die die Entfernung oder den Zugriff auf einzelne Beiträge betreffen, (i) Summierungseffekte, soweit ein Medium von mehreren Sperrungsverpflichtungen betroffen ist, (ii) der Umsetzungsaufwand, soweit – wie in der presserechtlichen Rechtsprechung üblich – über den konkret gemeldeten Inhalt hinaus Prüfung und ggf. Entfernung bzw. Zugriffsbeschränkung umgesetzt werden sollen (C. 2. d)).

Eine automatische Ersetzung, Sperrung oder Löschung über das gesamte Archiv scheidet wegen des hohen Aufwands für die erforderliche manuelle Nachkontrolle praktisch aus (C. 2. d)).

Eine Überprüfung der Altbeiträge vor dem Online-Stellen – oder gar eine kontinuierliche Überprüfung bereits veröffentlichter Beiträge – kommt aufgrund des damit verbundenen wirtschaftlichen Aufwands einem faktischen Online-Veröffentlichungsverbot gleich (C. 3.).

D. Anmerkungen zu den grundrechtsrelevanten Auswirkungen der angesprochenen Maßnahmen

Während das Entfernen eines vollständigen Beitrags aus einem Online-Archiv nicht nur dessen Auffindbarkeit über Suchmaschinen, sondern auch dessen Abrufbarkeit insgesamt verhindert, mithin der Zugang zu der gesamten Berichterstattung für jedermann unterbunden würde, wirken Maßnahmen, die den Zugriff von Suchmaschinen auf solche konkreten Beiträge verhindern oder einschränken, „nur“ auf dessen Auffindbarkeit über Suchmaschinen. Je nach technischer Ausgestaltung bleiben die Inhalte dabei allerdings u. U. noch indirekt über Suchmaschinen recherchierbar. Die Inhalte bleiben zudem in jedem Fall über eine direkte Eingabe der URL, die Nutzung der Archiv-Suchfunktion des konkreten Mediums oder über Verlinkungen von anderen Seiten weiter abrufbar. Bei entsprechender Ausgestaltung (X-Robots-Tag im HTTP-Header, HTML-Meta-Tag, Browser-Weiche) würden aber zumindest die – für den Beschwerdeführer offenbar besonders belastenden – prominent in der Trefferliste platzierten Ergebnisse zu seiner Person in den Suchmaschinen-Ergebnissen entfallen.

Bei Umsetzen des „minus“ zur vollständigen Entfernung eines betroffenen Beitrags, d.h. dem Anonymisieren bzw. Entfernen des Namens des Betroffenen, bliebe der Beitrag zum einen weiter abrufbar und zum anderen weiter über Suchmaschinen, nämlich etwa über die Eingabe anderer Suchworte, auffindbar.

Bei jeder Löschung oder Veränderung von Inhalten würden zugleich gleichlautende namentliche Recherchen über die Suchfunktion des jeweiligen Mediums verhindert. Sofern nicht das Medium zwei

unterschiedliche Archive führt, wäre davon auch die redaktionsinterne Recherche betroffen, was aus Sicht der DGRI zumindest eine erhebliche Einschränkung der Arbeit der Presse darstellen würde. Dies gilt besonders, wenn nicht einmal erkennbar wäre, dass ein Beitrag nachträglich verändert wurde, oder wenn bei einem gelöschten Beitrag der Anschein der Vollständigkeit der Trefferliste entstünde.

Soweit eine gerichtlich verfügte Einschränkung der Veröffentlichungsbefugnis sich nur auf konkret benannte Beiträge bezieht, sind erhebliche negative Auswirkungen auf die Pressearbeit höchstens aufgrund von Summierungseffekten wegen einer Vielzahl geringer Belastungen durch eine Vielzahl von Verboten denkbar.

Die übliche – im Ausgangsverfahren in den ersten beiden Instanzen allerdings auf das Nennen des Nachnamens des Beschwerdeführers beschränkte – umfassende Verbotstenorierung "jeglicher identifizierender Berichterstattung" stellt Betreiber von Online-Archiven letztlich vor schwer lösbare Probleme, weil grundsätzlich der gesamte Archivbestand eingehend geprüft werden müsste. Gleiches gilt für eine – vom Berufungsgericht im Ausgangsverfahren angenommene – Verpflichtung, sämtliche Berichterstattung zumindest vor dem Einstellen in ein Online-Archiv, wenn nicht gar regelmäßig, zu prüfen – das käme letztlich einem Verbot von Online-Archiven gleich.

Nach Ansicht der DGRI sind mithin Verbote kritisch zu sehen, die sich nicht nur auf konkret in einer Beschwerde benannte Beiträge, sondern etwa einen bestimmten Gegenstand der Berichterstattung beziehen. Eine graduell grundrechtsschonendere Umsetzung etwaiger Verbote könnte darin bestehen, dass der Betreiber des Online-Archivs für die konkret vom Betroffenen benannten Beiträge im HTML- oder HTTP-Header eine Anweisung an Suchmaschinen einfügt, den jeweiligen Beitrag nicht zu indizieren, und/oder – als zweitbeste, aber in der Praxis wohl ähnlich wirksame Lösung – eine Browser-Weiche einsetzt, die Suchmaschinen die betroffenen Inhalte nicht ausliefert.

Die Gesellschaft steht Ihnen für ergänzende Erläuterungen gerne zur Verfügung.

Mit freundlichen Grüßen

Dr. Anselm Brandi-Dohrn
Rechtsanwalt
Vorsitzender der DGRI e.V.