

Digitalwirtschaft 2030

Bernzen / Heinze / Steinrötter

2026

ISBN 978-3-406-85394-4

C.H.BECK

schnell und portofrei erhältlich bei
beck-shop.de

Die Online-Fachbuchhandlung beck-shop.de steht für Kompetenz aus Tradition. Sie gründet auf über 250 Jahre juristische Fachbuch-Erfahrung durch die Verlage C.H.BECK und Franz Vahlen.

beck-shop.de hält Fachinformationen in allen gängigen Medienformaten bereit: über 12 Millionen Bücher, eBooks, Loseblattwerke, Zeitschriften, DVDs, Online-Datenbanken und Seminare. Besonders geschätzt wird beck-shop.de für sein umfassendes Spezialsortiment im Bereich Recht, Steuern und Wirtschaft mit rund 700.000 lieferbaren Fachbuchtiteln.

§ 10 Maschinenlesbarkeit von Daten in Datenschutz und Urheberrecht

Prof. Dr. Ralf Kneuper, LL.M.

Zusammenfassung

In bestimmten Anwendungsfällen ist es aus Sicht des Gesetzgebers erforderlich, dass festgelegte Informationen von einer Partei einer anderen so bereitgestellt werden, dass der Empfänger diese elektronisch weiterverarbeiten kann. Dazu müssen die Informationen in einem maschinenlesbaren Format bereitgestellt werden. Dieser scheinbar eindeutige Begriff der Maschinenlesbarkeit kann allerdings in unterschiedlichen Kontexten sehr unterschiedliche Bedeutungen haben, so beispielsweise bei der Datenübertragbarkeit nach Art. 20 DS-GVO gegenüber dem TDM-Nutzungsvorbehalt nach § 44b UrhG. Die vorliegende Ausarbeitung untersucht die damit verbundenen unterschiedlichen Anforderungen an die Maschinenlesbarkeit von Informationen und stellt die wichtigsten technischen Lösungen zur Umsetzung vor. Insbesondere für TDM-Nutzungsvorbehalte wird dabei deutlich, dass die aktuell meist empfohlenen Umsetzungen wie die Nutzung einer robots.txt-Datei die relevanten Anforderungen nicht angemessen erfüllen.

A. Einführung und Motivation

Maschinenlesbarkeit ist eine gesetzliche Forderung, mit der der Gesetzgeber den Austausch von Informationen zwischen verschiedenen Parteien ermöglichen will. Dabei soll eine Partei festgelegte Informationen als Daten in einer solchen Form bereitstellen, dass der Empfänger diese Daten elektronisch weiterverarbeiten und die darin enthaltenen Informationen leicht wieder herleiten kann.¹ Beispiele dafür sind Art. 20 Abs. 1 DS-GVO (Datenübertragbarkeit) und § 44b Abs. 3 S. 2 UrhG (Nutzungsvorbehalt bei Text und Data Mining TDM). Auch wenn beide Forderungen nach Maschinenlesbarkeit vom Wortlaut her sehr ähnlich sind, handelt es sich technisch und inhaltlich doch um wesentlich unterschiedliche Forderungen.

Bei der Datenübertragbarkeit nach DS-GVO geht es um die Übertragung nennenswerter Datenmengen, wobei das verwendete Format sowie der verwendete Übertragungsweg vorher abgesprochen oder zumindest vom Sender benannt werden. Die Herausforderung besteht vor allem in der Auswahl eines geeigneten und für beide Seiten akzeptablen Datenformats.

Eine völlig andere Situation liegt beim TDM-Nutzungsvorbehalt vor. Hier geht es um eine geringe Datenmenge, ggf. sogar um nur ein Bit: „Nutzungsvorbehalt liegt vor/liegt nicht vor“. Die wesentliche Herausforderung besteht darin zu klären, an welcher Stelle und in welchem Format dieses Datum abgelegt wird, so dass potenzielle Nutzer den TDM-Nutzungsvorbehalt auch finden und als solchen erkennen.

¹ Kneuper DSRITB 2025, 1 ff.

- 4 Ein weiterer Unterschied zwischen beiden Formen der Maschinenlesbarkeit mit erheblichen praktischen Auswirkungen betrifft die damit verbundenen Interessen der Beteiligten. Bei der Datenübertragbarkeit hat der Verantwortliche als Sender der Daten grundsätzlich ein Interesse, die Daten mit geringem Aufwand, ohne Diskussionen oder Nachfragen, bereitzustellen, dh die Interessen beider Seiten in Bezug auf die Maschinenlesbarkeit sind weitgehend identisch. Der vom Rechteinhaber bereitgestellte TDM-Nutzungsvorbehalt bedeutet dagegen eine Einschränkung für den Empfänger (im Folgenden als TDM-Anwender bezeichnet), der damit ein Interesse hat, dass der Vorbehalt nicht auffindbar bzw. nicht maschinenlesbar und damit nicht gültig ist.
- 5 Aus diesen sehr unterschiedlichen Rahmenbedingungen, die sich aus der Forderung nach Maschinenlesbarkeit ergeben, folgen auch unterschiedliche Lösungsansätze. Ziel des vorliegenden Beitrags ist es, die technischen Möglichkeiten zur Umsetzung der Maschinenlesbarkeit in beiden Fällen kurz zu erläutern, um darauf aufbauend zu bewerten, inwieweit die jeweiligen rechtlichen Anforderungen damit erfüllt werden können.
- 6 Eine weitere Bedeutung des Begriffs der Maschinenlesbarkeit wurde in der KI-Verordnung eingeführt, wo gefordert wird, dass Ausgaben eines KI-Systems als künstlich erzeugt gekennzeichnet werden, wobei diese Kennzeichnung in einem maschinenlesbaren Format geschehen muss (Art. 50 Abs. 2 KI-VO). Diese Regelung ist derzeit noch nicht in Kraft² und betrifft viele unterschiedliche Dateiformate, wodurch auch unterschiedliche Kennzeichnungen erforderlich werden. Daher wird diese Form nach Maschinenlesbarkeit hier nicht weiter betrachtet.³ Ebenfalls nicht weiter betrachtet werden verwaltungsrechtliche Anforderungen, zB an die Maschinenlesbarkeit von Ausweisen. Der Fokus dieser Untersuchung liegt auf der Bedeutung des Begriffs der Maschinenlesbarkeit inkl. der eng verwandten Begriffe der gängigen, strukturierten und offenen Formate. Welche Daten unter welchen Voraussetzungen bereitgestellt werden müssen, inwieweit beispielsweise auch das Training von KI-Systemen unter die TDM-Schranke fällt, ist nicht Thema der Untersuchung.

B. Maschinenlesbarkeit im Datenschutz

- 7 Die Maschinenlesbarkeit von Daten wird in der DS-GVO an zwei Stellen gefordert, nämlich einerseits im Kontext des Rechts auf Datenübertragbarkeit (Art. 20 DS-GVO) sowie für elektronisch dargestellte Bildsymbole für Datenschutzinformationen (Art. 12 Abs. 7 DS-GVO). Beide Forderungen werden im Folgenden untersucht, mit Schwerpunkt auf die Maschinenlesbarkeit für die Datenübertragbarkeit.

I. Der Begriff der Maschinenlesbarkeit in Art. 20 DS-GVO

- 8 Gemäß Art. 20 Abs. 1 DS-GVO haben Betroffene das Recht, sie betreffende personenbezogene Daten unter bestimmten Voraussetzungen in einem strukturierten, gängigen und maschinenlesbaren Format zu erhalten. Technisch sehr ähnliche Forderungen sind ua in § 327p Abs. 3 BGB (Vertragsbeendigung bei digitalen Produkten) sowie Art. 3 Abs. 1, 4

² Ursprünglich wurde festgelegt, dass Art. 50 ab dem 2.8.2026 gelten soll. Zum Zeitpunkt der Erstellung dieses Beitrags ist diese Festlegung noch in Kraft, aber es wird eine Verschiebung um drei oder sechs Monate zwischen Parlament und Rat diskutiert.

³ Erste Vorschläge, wie Maschinenlesbarkeit in diesem Kontext zu verstehen ist, sind beispielsweise im „Second Draft Code of Practice on Transparency of AI-Generated Content“ der Europäischen Kommission enthalten.

Abs. 1 Data Act (Zugänglichmachen von Produkt- und Dienstdaten; auch über API) enthalten. In allen diesen Fällen geht es darum, nennenswerte Datenmengen an definierte Empfänger zu übertragen. Damit die Empfänger diese Daten angemessen auswerten und verarbeiten können, müssen die Daten in einem maschinenlesbaren Format bereitgestellt werden, wobei dieser Begriff in den genannten Gesetzen nicht legaldefiniert ist. In der Literatur wird daher meist auf die Definition in der PSI-RL 2013 verwiesen⁴ bzw. deren Nachfolger, Art. 2 Abs. 13 OD/PSI-RL, oder die deutsche Umsetzung im Datennutzungsgesetz (DNG). Laut § 3 Abs. 5 DNG „liegt ein maschinenlesbares Format vor, wenn die Daten durch Software automatisiert ausgelesen und verarbeitet werden können“.⁵ Etwas ausführlicher beschreibt Art. 2 Abs. 13 OD/PSI-RL ein maschinenlesbares Format als „Dateiformat, das so strukturiert ist, dass Softwareanwendungen konkrete Daten, einschließlich einzelner Sachverhaltsdarstellungen und deren interner Struktur, leicht identifizieren, erkennen und extrahieren können“. In Erwgr. 35 OD/PSI-RL wird ergänzend erläutert, dass ein maschinenlesbares Format offen oder proprietär sein kann und einem formellen Standard entsprechen kann oder nicht. Wenn die Daten nicht oder nicht ohne Weiteres aus einem Dokument extrahiert werden können, dann gilt das Dokument nicht als maschinenlesbar. In § 7 Abs. 2 DNG ist darüber hinaus gefordert, dass Daten soweit möglich und sinnvoll in offenen, maschinenlesbaren, zugänglichen, auffindbaren und interoperablen Formaten zusammen mit den zugehörigen Metadaten bereitzustellen sind. Damit stellt sich die Frage, wie die verschiedenen genannten Eigenschaften von Datenformaten zusammenhängen: Gemäß Erwgr. 35 OD/PSI-RL sind Offenheit und Standardkonformität Eigenschaften, die unabhängig von der Maschinenlesbarkeit eines Datenformats ist. Aus Art. 2 Abs. 13 OD/PSI-RL folgt dagegen, dass die Strukturiertheit der Daten ein erforderlicher Aspekt der Maschinenlesbarkeit ist. Interoperabilität ist in § 7 Abs. 2 DNG als separate Eigenschaft formuliert, aber aus technischer Sicht ist Interoperabilität nur erreichbar mit einem maschinenlesbaren Datenformat.⁶

Voraussetzung für die Auswertbarkeit und damit Maschinenlesbarkeit von Daten ist, dass eine Beschreibung der verwendeten Datenstruktur vorliegt, also die zugehörigen Metadaten,⁷ auch wenn das in der DS-GVO nicht ausdrücklich gefordert ist. Ohne die zugehörigen Metadaten ist nicht erkennbar, welche Bedeutung bereitgestellte Daten haben. Liegt zB wie unten beschrieben in einem Datensatz zu einer Person der Wert „Universität Potsdam“ vor, so ist daraus nicht erkennbar, ob es sich um den Studienort, den Arbeitgeber, den Ort einer besuchten Tagung oder eine andere Information handelt.

II. Geeignete Datenformate

1. Überblick

Als geeignete Datenformate werden meist XML, JSON und CSV genannt⁸, gelegentlich auch SQL. Gemeinsam ist diesen Formaten, dass sie jeweils aus verschiedenen Variablen, also Feldern mit einem Namen und einem Wert bestehen. (Bei CSV ist der Name optional und kann auch weggelassen werden.) Die Formate unterscheiden sich, neben den Details der

⁴ So zB in Art. 29 Datenschutzgruppe WP 242 S. 20; auch EU-Verordnungen wie zB 2023/2859, ABl. L, 2023/2859 und 2024/903, ABl. L, 2024/903, verwenden den Begriff der Maschinenlesbarkeit und verweisen dabei auf die Definition in Art. 2 Abs. 13 OD-PSI-RL, ABl. 2019 L 172, 56.

⁵ Fast wortgleich findet sich diese Definition auch in § 12 Abs. 1 S. 2 EGovG.

⁶ Art. 29 Datenschutzgruppe WP 242 S. 20; andeutungsweise auch in Hartl/Ludin/Werthschulte DNG Einleitung Rn. 22.

⁷ Art. 29 Datenschutzgruppe WP 242 S. 20; ähnlich Art. 5 Abs. 1 OD/PSI-RL sowie § 12 Abs. 1 S. 3 EGovG.

⁸ So zB Art. 29-Datenschutzgruppe, WP 242, 2017, 21.

reinen Darstellung, vor allem dadurch, inwieweit eine hierarchische Struktur beschrieben werden kann, und inwieweit die erlaubten Werte eingeschränkt werden können, beispielsweise auf Zahlen, Datumswerte oder Zeichenketten einer bestimmten Länge. Hauptzweck einer solchen Einschränkung ist es, Fehler durch ungültige Werte zu verhindern. Nicht geeignet sind nach hM textbasierte Datenformate wie PDF, Textverarbeitungsdateien oder HTML,⁹ da eine automatisierte Auswertung bei diesen Formaten zwar prinzipiell möglich, aber sehr aufwendig ist. Anders formuliert bedeutet das, dass Maschinenlesbarkeit keine binäre Eigenschaft ist, sondern es unterschiedliche Grade der Maschinenlesbarkeit gibt.

- 11 Selten ausdrücklich erwähnt wird, dass alle oben als geeignet genannten Formate noch sehr abstrakt sind und konkretisiert werden müssen durch Festlegung von Namen und Bedeutung der Tags (XML) bzw. Keys (JSON) bzw. Spaltennamen (Tabellenformate). Ohne diese Konkretisierung weiß man beispielsweise, dass bestimmte Daten im Feld „Ort“ stehen, aber nicht, welcher Ort (Geburtsort, Wohnort, Reiseziel etc.) gemeint ist. Darüber hinaus ist der Name eines Feldes auch nicht immer sprechend, sondern könnte beispielsweise auch „x42“ lauten. Grundsätzlich sollte bei XML, JSON und SQL mindestens das jeweils verwendete Schema (mit aussagekräftigen Feldnamen) bereitgestellt werden, bei CSV mindestens aussagekräftige Feldnamen als Überschriftenzeile. Besser ist ein zusätzliches Dokument, in dem die Bedeutung der Felder genau definiert wird. Soweit für bestimmte Typen von Anwendungen ein standardisiertes Austauschformat existiert, sollte dieses verwendet werden, auch wenn das derzeit eher die Ausnahme sein dürfte. Als Beispiel zur Erläuterung der verschiedenen Formate wird jeweils der folgende Sachverhalt betrachtet: Die zu übertragenden Informationen umfassen die Tatsache, dass eine Person an der DSRI-Herbstakademie 2026 teilnimmt, die vom 9.9. bis 12.9.2026 an der Universität Potsdam stattfindet.

2. XML

- 12 Die eXtensible Markup Language (XML) ist eine Sprache zur Beschreibung komplexer Datenstrukturen und insbesondere als Datenaustauschformat geeignet. Sie ermöglicht die Beschreibung sehr strukturierter Formate, ist dadurch aber komplex und führt zu einem relativ großen Speicherplatzbedarf. Der oben genannte Sachverhalt kann in XML beispielsweise wie folgt beschrieben werden:

```
<tagung>
<name>DSRI-Herbstakademie 2026</name>
<ort>Universität Potsdam</ort>
<datum>
<start>2026-09-09</start>
<ende>2026-09-12</ende>
</datum>
</tagung>
```

- 13 XML definiert die hier verwendete hierarchische Datenstruktur, ist alleine aber nicht hinreichend konkret, da die verwendeten (als „Tags“ bezeichneten) Kennzeichnungen wie zB <tagung> ... </tagung> anwendungsspezifisch und daher nicht Teil von XML sind. Diese anwendungsspezifischen Metadaten werden ihrerseits als Schema in Formaten wie XML Schema oder RelaxNG definiert und legen neben dem Namen von Tags weitere

⁹ Hartl/Ludin/Werthschulte DNG/Hartl/Ludin § 3 Rn. 35; Keuchen/Deuber RD i 2022, 189 Rn. 19; Art. 29-Datenschutzgruppe, WP 242, 21; aA Richter DNG § 3 Rn. 214; BeckOK InfoMedienR/Debus DNG § 3 Rn. 38.

Rahmenbedingungen fest, beispielsweise ob ein bestimmter Tag verpflichtend oder optional ist und welche Datentypen (Zeichenkette mit einer bestimmten (Maximal-)Länge, Zahl etc) für die jeweiligen Werte erlaubt sind. Für viele Anwendungsfälle gibt es standardisierte XML-Schemata als Austauschformate, beispielsweise für die e-Rechnung oder als XÖV-Standard für die öffentliche Verwaltung. Für den hier betrachteten Anwendungsfall der Datenübertragung von personenbezogenen Daten nach Art. 20 DSGVO sind solche standardisierten Formate aber kaum vorhanden.

3. JSON

Die JavaScript Object Notation (JSON) ist ein konzeptionell ähnliches aber sehr viel einfacheres Datenformat, das für den Datenaustausch zwischen Servern und Webapplikationen entwickelt wurde. Es handelt sich dabei um eine einfache Textform, die gut lesbar ist für Menschen (vorausgesetzt es werden sprechende Feldnamen verwendet) wie auch Maschinen. Im Wesentlichen werden Daten als Objekte dargestellt, die aus einer Menge von Paaren von Schlüsseln (Namen) und Werten (Key-Value-Paare) bestehen. In JSON lässt sich der obige Sachverhalt wie folgt darstellen:

```
{„Tagungsname“:„DSRI-Herbstakademie 2026“,  
  „Ort“: „Universität Potsdam“,  
  „Datum“:  
  {„Startdatum“: „2026-09-09“,  
    „Enddatum“: „2026-09-12“  
  }  
}
```

Strukturell sieht das sehr ähnlich aus wie der entsprechende Text in XML, da bei diesem Beispiel auf die Nutzung der vielen zusätzlichen XML-Strukturelemente verzichtet wurde.

Analog XML benötigt man auch hier ein Schema zur Definition der verwendeten Struktur (Metadaten), wobei die dafür verwendete Sprache JSON Schema sich auf die Festlegung von Namen und einfacher Datentypen beschränkt. Hier gibt es kaum standardisierte Formate, so dass man das verwendete Schema einschließlich der (über den Namen hinausgehenden) Bedeutung der genutzten Felder meist individuell vereinbaren muss, um JSON-Daten austauschen zu können.

Allerdings gibt es mit JSON-LD eine Erweiterung des JSON-Formats, die die Einbindung von standardisierten Definitionen (soweit vorhanden) der verwendeten Schlüssel, also der Metadaten, ermöglicht:

```
„@context“: „https://schema.org“,  
„@type“: „Event“,  
„name“: „DSRI-Herbstakademie 2026“,  
„location“: {  
  „@type“: „Place“,  
  „address“: {  
    „@type“: „PostalAddress“,  
    „addressLocality“: „Potsdam“  
  },  
  „startDate“: „2026-09-29“,  
  „endDate“: „2026-09-12“  
}
```

4. CSV

- 17 Das Format Comma-Separated Value (CSV)¹⁰ ist ein sehr einfaches Tabellenformat, in dem ein Datensatz jeweils als eine Zeile dargestellt wird, die aus einer Liste von Werten besteht. Im CSV-Format kann der obige Sachverhalt daher wie folgt dargestellt werden:

```
Tagungsname, Ort, Startdatum, Enddatum
DSRI-Herbstakademie 2026, Universität Potsdam, 2026-09-09, 2026-09-12
```

Bei CSV-Format beschränken sich die Metadaten meist auf eine Überschriftenzeile zu Beginn der Tabelle, und sogar auf diese wird manchmal verzichtet. Optional, aber selten genutzt und ohne standardisiertes Format, gibt es eine separate Beschreibung, was die einzelnen Feldnamen bedeuten.

5. SQL

- 18 SQL¹¹ nutzt ähnlich CSV ein Tabellenformat mit Zeilen, die ihrerseits aus Feldern bestehen. Eine hierarchische Struktur der Daten, wie oben bei XML oder JSON für das Datum verwendet, ist bei SQL also nur indirekt möglich, indem die Daten auf mehrere Tabellen aufgeteilt werden und aus einem Feld auf einen Datensatz einer anderen Tabelle verwiesen wird. (Eine solche Aufspaltung ist prinzipiell zwar auch bei CSV möglich, wird bei diesem sehr einfachen Format aber selten verwendet.) SQL ist eine zentrale technische Grundlage der (relationalen) Datenbanken und enthält Befehle zur Definition und Verwaltung von Datenbanken sowie der darin enthaltenen Daten. Im vorliegenden Zusammenhang ist vor allem der INSERT-Befehl relevant, mit dem Daten in eine Datenbank-tabelle geschrieben werden können und der daher auch als Grundlage für den Datenaustausch dienen kann. Der obige Sachverhalt kann in SQL beispielsweise wie folgt beschrieben werden:

```
INSERT INTO Tagung (Name, Ort, Startdatum, Enddatum) VALUES ('DSRI-Herbstaka-
demie 2026', 'Universität Potsdam', 2026-09-09, 2026-09-12);
```

Die zu verwendenden Felder werden bei SQL durch die Tabellenstruktur festgelegt, für die es einen eigenen Befehl „CREATE TABLE“ gibt. Mit diesem SQL-Befehl werden die relevanten Metadaten vorgegeben, also Feldnamen, zugehörige Datentypen und Festlegungen, ob Felder gefüllt sein müssen.

III. Maschinenlesbarkeit von Bildsymbolen

- 19 Neben der bisher beschriebenen Maschinenlesbarkeit im Kontext des Rechts auf Datenübertragbarkeit wird Maschinenlesbarkeit in der DS-GVO auch für elektronisch dargestellte Bildsymbole für Datenschutzinformationen gefordert (Art. 12 Abs. 7 S. 2 DS-

¹⁰ Komma-getrennte Werte, dh ein CSV-Datensatz besteht meist aus einer Liste von Werten, die durch Kommas voneinander getrennt werden. Je nach Kontext werden als Trennzeichen aber oft auch andere Zeichen wie das Semikolon verwendet, auch wenn in diesem Fall die Bezeichnung CSV nicht mehr ganz korrekt ist.

¹¹ SQL wurde ursprünglich als Structured Query Language beschrieben, aber heute als eigenständiger Name und nicht als Akronym verstanden.

GVO).¹² In erster Linie sind diese Symbole für menschliche Leser gedacht, aber zusätzlich ist Maschinenlesbarkeit gefordert, wenn die Bildsymbole in elektronischer Form dargestellt werden. Zweck dieser Forderung ist es, die Automatisierung von Betroffenenrechten zu erleichtern.¹³

Auf einer sehr oberflächlichen Ebene sind elektronisch dargestellte Symbole natürlich immer maschinenlesbar. Die Forderung nach Maschinenlesbarkeit bezieht sich darauf, dass die Bedeutung der Bildsymbole leicht von Software erkannt und identifiziert werden kann. Dies ist beispielsweise nicht der Fall, wenn die Bildsymbole nur als Grafiken (zB im JPG-, PNG- oder BMP-Format) vorliegen. Maschinenlesbarkeit kann in diesem Fall vor allem durch die Nutzung von Alternativtexten für die Grafiken erreicht werden. Im für Web-Seiten verwendeten HTML-Format kann eine Grafik mit Alternativtext beispielsweise wie folgt dargestellt werden:

```
<img src=„https://www.baden-wuerttemberg.datenschutz.de/wp-content/uploads/daten  
schutzicons/katalog/1.%20Verantwortliche_r/SW_1.0%20Verantwortliche_r.png#„ alt=„Ver-  
antwortliche_r“>
```

Ein solcher Alternativtext und die damit verbundene Maschinenlesbarkeit ist auch erforderlich zur Barrierefreiheit, um beispielsweise den Inhalt einer Grafik für Seheingeschränkte vorlesen zu können.

Auch bei Verwendung von Alternativtexten bleibt allerdings die Schwierigkeit, dass diese Texte nicht standardisiert sind, wodurch die automatisierte Auswertung und Verarbeitung wesentlich erschwert werden.

IV. Zwischenfazit

Die Herausforderung besteht in diesem Zusammenhang nicht darin, ob ein für die Bereitstellung von Daten gemäß Art. 20 DS-GVO verwendetes Format als maschinenlesbar im engeren Sinn bewertet wird, sondern ob es auch ausreichend strukturiert und gängig ist. Eine solche Strukturierung schließt ein, dass auch die zugehörigen Metadaten in ausreichendem Detailgrad bereitgestellt werden, damit die Daten interpretiert und ausgewertet werden können. Die Forderung nach der Maschinenlesbarkeit von Bildsymbolen für Datenschutzinformationen hat derzeit eine sehr geringe praktische Bedeutung, ist andererseits aber relativ einfach umzusetzen durch Nutzung von Alternativtexten.

C. Maschinenlesbarkeit von Nutzungsvorbehalten im Text und Data Mining

I. Überblick

Nach § 44b UrhG, der deutschen Umsetzung von Art. 4 Abs. 3 DSM-RL, ist die TDM-Analyse von urheberrechtlich geschützten digitalen oder digitalisierten Werken grundsätzlich erlaubt, sofern der Rechtsinhaber sich diese Nutzung nicht vorbehalten hat. Für online zugängliche Werke muss ein solcher Nutzungsvorbehalt in maschinenlesbarer

¹² Geeignete standardisierte Bildsymbole (Icons) wurden beispielsweise vom LfDI BW sowie vom Bitkom veröffentlicht, haben sich aber bislang nicht durchgesetzt.

¹³ Kühling/Buchner DS-GVO BDSG/Bäcker Art. 12 Rn. 22.

Form erfolgen (§ 44b Abs. 3 S. 2 UrhG), um eine automatisierte Auswertung zu ermöglichen.¹⁴

25 Um die Maschinenlesbarkeit eines Nutzungsvorbehalts zu analysieren, muss geklärt sein, welche Informationen er enthalten muss:

- Auf welches Werk bzw. welche Werke bezieht sich der Nutzungsvorbehalt? Meist soll der Nutzungsvorbehalt für alle Inhalte einer Webpräsenz gelten, aber auch eine Beschränkung auf ausgewählte Inhalte ist möglich, beispielsweise alle eingebundenen Fotos oder auch nur eine einzelne Seite.
- Auf welche Nutzungen bezieht sich der Vorbehalt? Der Nutzungsvorbehalt kann sich auf alle TDM-Nutzungen beziehen, aber auch eine Einschränkung ist möglich. Der GPAI Code of Practice der EU¹⁵ beispielsweise empfiehlt ausdrücklich eine Unterscheidung zwischen Nutzungseinschränkungen in Bezug auf die Indexierung für Suchmaschinen und auf das Training von GPAI-Modellen. Ohne die Möglichkeit zu einer solchen Unterscheidung würde ein Vorbehalt in Bezug auf das Training von GPAI-Modellen dazu führen, dass die betroffenen Inhalte nicht mehr in Suchmaschinen aufzufinden wären. Denkbar ist auch, dass ein Rechteinhaber beim Nutzungsvorbehalt nach (Gruppen von) TDM-Nutzern unterscheiden möchte und beispielsweise eine Nutzung durch akademische Organisationen erlaubt, sich die gleiche Nutzung durch kommerzielle Organisationen aber vorbehält.
- Falls nicht aus dem Kontext erkennbar, muss auch der Rechteinhaber, der sich die Nutzung vorbehält, entsprechend benannt werden.

Ein allgemeingültiges Format für die Darstellung maschinenlesbarer Nutzungsvorbehalte muss die Formulierung all dieser Informationen ermöglichen. Zusätzlich dazu muss in diesem Fall auch der Ort der Speicherung als Aspekt der Maschinenlesbarkeit berücksichtigt werden, der im Gesetz offengelassen ist. Im Folgenden werden die verbreitetsten Vorschläge für Format und Speicherort maschinenlesbarer TDM-Nutzungsvorbehalte zusammengefasst und erläutert, insbesondere das Robot Exclusion Protocol (robots.txt), das Impressum oder die AGB sowie das TDM Reservation Protocol.

II. Datenformate

1. Das Robot Exclusion Protocol (robots.txt)

26 Das Robot Exclusion Protocol (REP) wurde 1994 ursprünglich für Suchmaschinen-Robots (in der Protokolldefinition als Crawler oder User Agents bezeichnet) formuliert und 2022 als Standard RFC 9309 der Internet Engineering Task Force (IETF) veröffentlicht. Gemäß diesem Protokoll werden Vorgaben und Einschränkungen für Robots als Gruppen von Regeln in einer Datei robots.txt beschrieben, die im Wurzelverzeichnis eines Webservers abgelegt wird. Jede Gruppe besteht aus einem oder mehreren Robots sowie einer oder mehreren Allow- bzw. Disallow-Regeln, die festlegen, auf welche Pfade und Dateien die Robots (nicht) zugreifen dürfen. Ein Beispiel:

```
User-Agent: GPTBot
Disallow:/
Allow: /public/
```

¹⁴ BT-Drs. 19–27426, 89.

¹⁵ Copyright Chapter, Measure 1.3, Nr. 5.