

Maschinenlesbarkeit von Daten in Datenschutz und Urheberrecht

Prof. Dr. Ralf Kneuper, LL.M.

IU Internationale Hochschule

Herbstakademie 2026

Die Forderung nach Maschinenlesbarkeit

- ▶ Ziel: Austausch von Informationen zwischen verschiedenen Parteien ermöglichen
- ▶ Beispiele
 - ▶ Art. 20 (1) DS-GVO (Datenübertragbarkeit)
 - ▶ § 44b (3) S. 2 UrhG (Nutzungsvorbehalt bei Text und Data Mining TDM)
 - ▶ (Art. 50 (2) KI-VO –Transparenz-Kennzeichen)
 - ▶ etc.
- ▶ Sehr ähnlicher Wortlaut, aber technisch-inhaltlich sehr unterschiedliche Forderungen

Maschinenlesbarkeit

Art. 20 (1) DS-GVO (Datenübertragbarkeit)

- ▶ Nennenswerte Datenmengen
- ▶ Format-Absprache zwischen Sender und Empfänger möglich
- ▶ Herausforderung:
 - ▶ Auswahl eines für beide Seiten akzeptablen, ausdrucksfähigen Formats

§ 44b (3) S. 2 UrhG (TDM-Nutzungsvorbehalt)

- ▶ Sehr geringe Datenmenge
- ▶ Widerstrebende Interessen von Sender und Empfänger
- ▶ Herausforderung:
 - ▶ Wo ablegen, damit Empfänger die Daten findet und erkennt

Art. 20 (1) DS-GVO (Datenübertragbarkeit)

- ▶ Die betroffene Person hat das Recht, die sie betreffenden personenbezogenen Daten, die sie einem Verantwortlichen bereitgestellt hat, in einem strukturierten, gängigen und **maschinenlesbaren** Format zu erhalten ...
- ▶ Inhaltlich ähnliche Anforderungen ohne Konkretisierung in BGB, Data Act etc.
- ▶ Oft Verweis auf Art. 2 (13) OD-PSI-RL
 - ▶ "Dateiformat, das so strukturiert ist, dass Softwareanwendungen konkrete Daten, einschließlich einzelner Sachverhaltsdarstellungen und deren interner Struktur, leicht identifizieren, erkennen und extrahieren können"
- ▶ Umgesetzt in § 3 (5) Datennutzungsgesetz (DNG):
 - ▶ "liegt ein maschinenlesbares Format vor, wenn die Daten durch Software automatisiert ausgelesen und verarbeitet werden können".

Maschinenlesbare Formate

Häufig genannt:

- ▶ XML
- ▶ JSON
- ▶ CSV
- ▶ (SQL)

Gemeinsame Eigenschaften dieser Formate:

- ▶ Bestehen aus Variablen, also Feldern mit Namen und Wert (und Typ)
- ▶ Zwei Ebenen:
 - ▶ Definition der Feldnamen und –typen (Metadaten)
 - ▶ Definition der Feldinhalte / Werte

eXtensible Markup Language (XML)

<tagung>

<name>DSRI-Herbstakademie 2026</name>

<ort>Universität Potsdam</ort>

<datum>

<start>2026-09-09</start>

<ende>2026-09-12</ende>

</datum>

</tagung>

- ▶ Hierarchische Datenstruktur
- ▶ Definition der Metadaten als Schema (XML Schema, RelaxNG)
- ▶ Beispiel-Schemata: e-Rechnung; XÖV-Standard für die öffentliche Verwaltung

JavaScript Object Notation (JSON)

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

- ▶ Konzeptionell ähnlich XML aber deutlich einfacher
 - ▶ XML ermöglicht komplexere Strukturen, Datentypen etc.
- ▶ Besteht i.W. aus Schlüssel-Wert-Paaren
- ▶ Erweiterung JSON-LD ermöglicht Einbindung standardisierter Strukturdefinition (Metadaten)

Comma-Separated Values (CSV)

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

- ▶ Überschriftenzeile optional, aber meist wichtig
- ▶ Tabellenformat, keine hierarchische Struktur möglich

SQL

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

- ▶ Tabellenformat, hierarchische Struktur nur indirekt möglich (durch Aufteilung von Daten auf mehrere Tabellen)
- ▶ Grundlage für (relationale) Datenbanken
- ▶ Struktur der Tabellen wird durch SQL-Befehl CREATE TABLE definiert und dann z.B. bei INSERT genutzt

Zwischenfazit Datenportabilität

- ▶ Herausforderung: Ist das verwendete Format ausreichend strukturiert und gängig?
 - ▶ Interpretierbarkeit Daten → Informationen
- ▶ Technische Lesbarkeit meist unproblematisch
- ▶ Erfordert Bereitstellung von Metadaten

§ 44b (3) UrhG (TDM-Nutzungsvorbehalt)

- ▶ Nutzungen nach Absatz 2 Satz 1 sind nur zulässig, wenn der Rechtsinhaber sich diese nicht vorbehalten hat. Ein Nutzungsvorbehalt bei online zugänglichen Werken ist nur dann wirksam, wenn er in **maschinenlesbarer** Form erfolgt.

Benötigte Informationen

- ▶ Nutzungsvorbehalt liegt vor? ja/nein
- ▶ Welche Werke?
- ▶ Rechteinhaber
- ▶ Optional:
 - ▶ Welche Nutzungen?
 - ▶ Indexierung in Suchmaschinen
 - ▶ GPAI-Training
 - ▶ ...
 - ▶ Welche Nutzer?
 - ▶ kommerziell
 - ▶ Forschung
 - ▶ ...

Robot Exclusion Protocol (robots.txt)

User-Agent: GPTBot

Disallow: /

Allow: /public/

- ▶ Ablageort klar definiert (Wurzelverzeichnis)
- ▶ Ermöglicht Einschränkungen in Bezug auf den Namen des Robots und der betroffenen Dateien und Verzeichnisse
 - ▶ Name muss bekannt sein; alternativ Einschränkung für alle Robots
- ▶ Keine Einschränkung in Bezug auf Art der Nutzung möglich
- ▶ Diverse Erweiterungen veröffentlicht (z.B. weboptout), aber nicht standardisiert
 - ▶ Eingeschränkt maschinenlesbar

Robot Exclusion Protocol (robots.txt)

- ▶ Wird in der juristischen Literatur zum Nutzungsvorbehalt oft empfohlen
- ▶ Verbindlichkeit umstritten
- ▶ Von Betreibern von Robots aber oft ignoriert

- ▶ Heise Verlag 2025:
 - ▶ *Nachruf: Abschied von robots.txt (1994-2025)*

Dokumentation in Impressum oder AGB

- ▶ Laut Gesetzesbegründung möglich, **sofern maschinenlesbar** (UrhDaG; BT-Drs. 19/27426, 89)
- ▶ Erweiterter Begriff der Maschinenlesbarkeit, der Auswertung durch KI einschließt?
 - ▶ So auch LG Hamburg 2024. Vom OLG Hamburg 2025 in Frage gestellt, aber mit Bezug auf den technischen Stand von 2021
 - ▶ aM: Würde die Forderung nach Maschinenlesbarkeit überflüssig machen
 - ▶ Denkbar, aber nicht überzeugend: Unterschiedliche Anforderungen an Maschinenlesbarkeit für KI-Anwendungen und andere Anwendungen

Dokumentation in Impressum oder AGB

- ▶ Überzeugend für einfache Formulierungen („Alle Rechte vorbehalten“) oder die Tatsache, dass Nutzungsvorbehalt vorliegt, aber nicht für Inhalte komplexerer Nutzungsvorbehalte
- ▶ Ablageort meist klar, erfordert aber Suche durch Nutzer

TDM (Text and Data Mining) Reservation Protocol

- ▶ Bericht (kein Standard) des W3C
- ▶ Beschreibt Spezifikation, die ausdrücklich als technische Antwort auf Art. 4 DSM-RL gedacht ist.

"tdm-reservation": "1",

"tdm-policy": "tdmrep.json"

- ▶ TDMRep beschreibt mehrere Optionen, wo diese Angaben abgelegt werden können
 - ▶ empfohlen: in der HTTP-Antwort des Webservers
- ▶ TDM-Policy erlaubt Vokabular für Formulierung von Nutzungsvorbehalten inkl. Angaben zu finanzieller Entschädigung
 - ▶ Erweiterungen in Arbeit

TDM Reservation Protocol

- ▶ Nicht breit bekannt, aber zumindest bei Verlagen in Deutschland und Frankreich bereits häufig eingesetzt
- ▶ Vokabular insb. für Nutzungsarten noch sehr eingeschränkt, aber Erweiterungen in Arbeit

Zwischenfazit

- ▶ Verbreitete Formate sind wenig ausdrucksstark (robots.txt) bzw. eingeschränkt maschinenlesbar (Impressum, AGB) und oft von TDM-Nutzern nicht als verbindlich akzeptiert
- ▶ Viele weitere Formate vorgeschlagen (über das hier vorgestellte TDMRep hinaus), sind aber wenig bekannt und Nutzungsvorbehalte daher schwer auffindbar
- ▶ Für TDM-Nutzer bedeutet das:
 - ▶ Suche nach Nutzungsvorbehalt an vielen Stellen erforderlich
- ▶ EU-Kommission hat ab 12/2025 Konsultation zu Protokollen für die praktische Umsetzung von Nutzungsvorbehalten durchgeführt
 - ▶ Ergebnisse und Konsequenzen sind bislang nicht veröffentlicht

Fazit

- ▶ Forderung nach Maschinenlesbarkeit kann zu sehr unterschiedlichen technischen Anforderungen führen
- ▶ Datenportabilität:
 - ▶ Viele mögliche Formate, aber ähnliche Interessen von Sender und Empfänger
 - ▶ Abstimmung nötig aber nicht besonders schwierig
- ▶ Nutzungsvorbehalt:
 - ▶ Viele mögliche Formate
 - ▶ Zentrale Herausforderungen sind Ausdrucksfähigkeit und Auffindbarkeit
 - ▶ Widerstrebende Interessen
 - ▶ Derzeit keine breit akzeptierten und aussagekräftigen Formate