

KI-Optionen im ONIX und Nutzungsvorbehalte digitaler Inhalte

Version 1.1: 27.11.2024



KI-Optionen im ONIX und Nutzungsvorbehalte digitaler Inhalte

Inhalt

Einleitung	3
1. Nutzungsvorbehalt	3
1. Ziel	3
1.1 Herausforderungen	3
1.2 Schutz im ONIX	4
1.2.1 Schutz des mitgelieferten digitalen Produkts	5
1.2.2 Schutz eines Marketingtextes	6
1.2.3 Schutz eines über einen Link zur Verfügung gestellten Inhalts	7
1.2.4. Schutz eines Covers	8
1.3 Schutz außerhalb von ONIX	9
1.3.1 Allgemeiner Schutz im Impressum	9
1.3.2 Schutz am digitalen Produkt	9
1.3.2.1 EPUB	9
1.3.2.2 Webseiten und HTML-Darstellungen	10
1.3.2.3 PDF	11
1.3.2.4 Cover (JPG)	11
1.4 Nutzungsbedingungen	12
1.5 Veranschaulichung der KI-Opt-out-Implementierungen	16
1.6 Fazit Nutzungsvorbehalt	16
2. Kennzeichnung von KI-Inhalten	17
2.1 KI als Contributor / Urheber*in	17
2.1.1 Covergestaltung	17
2.1.2. Übersetzung	19
2.2 KI-basierte Sprachausgabe in Hörbüchern	19
2.2.1 Personalisierte KI-Stimme	19
2.2.2 Nicht personalisierte KI-Stimme	20
2.3 Fazit Kennzeichnung	21

Einleitung

Systeme Künstlicher Intelligenz (KI), mit denen sowohl analytisch als auch generativ gearbeitet werden kann, sind bereits vielfach im Einsatz. Die Vorgaben des § 44b UrhG erfordern technische Umsetzungen für den Urheberrechtsschutz. Der AI-Act sieht vor allem im Bildbereich Kennzeichnungspflichten von KI generiertem Content vor. Zudem sollten freiwillige Kennzeichnungen standardisiert werden. Dieses Best Practice soll die Umsetzungsmöglichkeiten im ONIX und darüber hinaus aufzeigen.

1. Nutzungsvorbehalt

Die Basis künstlicher Intelligenz ist das Muster der Trainingsdaten. Damit sind die Trainingsdaten das Kapital der KI. Zum Trainieren einer KI sollen nur dafür freigegebene Daten genutzt werden. Inhalte dürfen nur dann genutzt werden, wenn die Urheber*innen sie für den Geschäftszweck der KI freigegeben haben (siehe auch <https://www.boersenverein.de/beratung-service/recht/kuenstliche-intelligenz/> sowie <https://www.boersenverein.de/politik-positionen/kuenstliche-intelligenz/>).

1. Ziel

Laut § 44b Abs. 3 UrhG ist bei online zugänglichen Werken ein Opt-out nur dann wirksam, wenn es in maschinenlesbarer Form erfolgt.

Dieses Kapitel beschreibt, wie Inhalte im digitalen Zwischenbuchhandel von der Verarbeitung von KIs und/oder dem sogenannten Text & Data Mining ausgenommen werden können. Dabei soll berücksichtigt werden, dass der Nutzungsvorbehalt als Kennzeichen möglichst nah am Content platziert wird.

Es soll ein abgestimmter Standard im Sinne eines Best Practice beschrieben werden, sodass die relevante Information mit den Inhalten so in den Markt gegeben werden kann, dass jede verarbeitende Stelle (insbesondere KIs) automatisch erkennen kann, ob ein Inhalt freigegeben ist oder nicht.

1.1 Herausforderungen

Es wird eher selten so sein, dass ein Inhalt **generell** von Text und Data Mining bzw. der Verarbeitung von KIs ausgenommen sein soll, gibt es doch KIs, die für Marketing oder Verkaufsförderung Inhalte analysieren – und gerade die Nutzung dieser Technologie ist in der Regel im Interesse der Urheber und Urheberinnen.

Mit „Inhalten“ können unterschiedlichste digitale Daten gemeint sein, z.B. ein EPUB oder PDF, eine Ausgabe auf einer Website (HTML), ein digitales Hörbuch (z.B. MP3), ein Cover (z.B. JPG), aber auch ein Klappentext oder eine Leseprobe.

Es gilt, technische Umsetzungen mit unterschiedlichen Freigaben für die unterschiedlichen Produkte zu definieren.

Es gibt aufgrund der Aktualität des Themas noch keinen offiziellen und verabschiedeten Web-Standard für den Ausschluss von Text- und Data Mining, sondern nur einen Vorschlag der Text

and Data Mining Reservation Protocol Community Group des W3C2, die sich aus Vertreter*innen unterschiedlichster Unternehmen u.a. aus dem Publishing-Bereich³ zusammensetzt.

Angesichts des Status des W3C als Standardisierungsgremium für das Web ist dies der derzeit vielversprechendste allgemeine Ansatz.

Daneben sind individuelle Opt-outs für einzelne Plattformen jederzeit möglich, z.B. nur für das Mining von Web-Inhalten oder Bildern.

1.2 Schutz im ONIX

Marktteilnehmer erhalten im Zusammenhang mit Assets bzw. Inhalten in der Regel ONIX-Daten. Mit Copyright-Angaben im ONIX ist eigentlich die Nutzung gelieferter Assets definiert. Das Best Practice "LICENCE TO USE EDItEUR STANDARDS" ([10.4400/nwgj](#)) führt dies aus:

- Bestimmte Daten, die in einer ONIX-Nachricht enthalten, zitiert oder verlinkt sind – z. B. Auszüge aus dem Produkt, Inhaltsverzeichnisse, beschreibende Texte, Cover und andere Ressourcen – unterliegen dem Urheberrecht (4.1).
- Wenn ein Datenlieferant einem Datenempfänger jedoch eine ONIX-Nachricht zur Verfügung stellt, besteht – auch ohne ausdrückliche Lizenz – eine klare Aufforderung an den Empfänger mit einer "stillschweigenden Lizenz", die bereitgestellten Produktdaten zum Zwecke der Katalogisierung, des Handels, des Merchandising, der Werbung und des Verkaufs der beschriebenen Produkte zu verwenden, sowohl innerhalb der Empfängerorganisation als auch in kundenorientierten Anwendungen (5.2).

Möglicherweise ist diese Information nicht allgemein bekannt oder auch zu unspezifisch. Daher trägt die Möglichkeit, explizite KI Opt-out Informationen zu mitgelieferten Assets im ONIX zu hinterlegen, zur besseren Darstellung und Auswertung bei.

TDM-Information sind derzeit jedoch noch nicht auf Assetebene ONIX verfügbar.

Es wird angestrebt, mit EDItEUR eine Lösung zu erarbeiten mit dem Ziel, dass Marktteilnehmer auf Basis der Information im ONIX handeln.

Einige Anwendungsbeispiele sollen verdeutlichen, welche Ziele ein Opt-out im ONIX verfolgt:

- a) Der Inhalt eines Full-EPUBs soll nicht von KIs verarbeitet werden, um z.B. eigene Texte daraus zu generieren (TDM-Reservation=1, ohne Policy)
- b) Der Inhalt eines Full-PDFs soll nicht von KIs verarbeitet werden, um z.B. eigene Texte daraus zu generieren (TDM-Reservation=1, ohne Policy)
- c) Der Inhalt einer mitgelieferten Leseprobe darf von KIs verarbeitet werden (TDM-Reservation=0)
- d) Cover werden Handelspartnern zur Verfügung gestellt, um die Produkte zu bewerben, jedoch nicht, um sie zum Trainieren einer Bild-KI zu nutzen. Damit wird verhindert, dass z.B. Selfpublishern des Portals aus Bestellern KI-generierte Cover angeboten werden (TDM-Reservation=1, ohne Policy)
- e) Cover sollen nicht zum Verarbeiten durch KIs genutzt werden dürfen, aber für die Generierung von Lesemotiven ist die Verarbeitung erlaubt. Diese Ausnahme wird in der Policy beschrieben (TDM-Reservation=1, TDM-Policy=<https://Musterverlag.de/richtlinien/richtlinie.json>)
- f) Handelspartner sollen darauf hingewiesen werden, welche Daten (von Verlagen) in den Portalen vor der KI-Verarbeitung geschützt werden. Die Handelspartner mögen die Information in der tdmrep.json (siehe [Kapitel 1.4.2](#)) hinterlegen. Die Information kann aus dem ONIX ausgelesen werden und kann sich auf alle mitgelieferten Assets beziehen (Cover, Leseprobe, PDF, EPUB, Hörproben etc.).

Ein Opt-Out für Print-Produkte im ONIX ist derzeit zwar technisch möglich aber fachlich nicht korrekt. Die im ONIX vorgesehene Sperre im EpubUsageConstraint bezieht sich (laut EDItEUR) nur auf E-Publications. Zudem sei die Digitalisierung und anschließende Weiterverarbeitung eines Printproduktes bereits über das bestehende Urheberrecht, bzw. Vervielfältigungsrecht, sowie der Angabe im Impressum ([siehe 1.4.1](#)) abgedeckt.

Das Einbetten der Kennzeichen liegt bei allen, die ONIX erstellen, das Auslesen bei allen, die KI einsetzen.

Mit der ONIX Version 3.1.1 kann ein Opt-out auf mehreren Ebenen erfolgen:

- Inhalt des mitgelieferten Werkes (EpubUsageConstraint auf Ebene DescriptiveDetail)
- Inhalte unter TextContent für jede Art des im ONIX gelieferten Textes (Beschreibung (03), Abstract (30).
- Inhalte unter SupportingResource für jedes Asset (einzeln), z.B. einem Coverlink

1.2.1 Schutz des mitgelieferten digitalen Produkts

Schon heute ist die Kennzeichnung eines Opt-Outs für den gelieferten Content der Formate EPUB, PDF und MP3 möglich.

Verwendet wird dafür das EpubUsageConstraint Composite (in Block 1 / ProductForm):

Abschnitt	ONIX-Tags	
	<EpubUsageConstraint>	
P.3.17	<EpubUsageType>11</EpubUsageType>	11 = Text and data mining
P.3.18	<EpubUsageStatus>03</EpubUsageStatus>	03 = Prohibited
	</EpubUsageConstraint>	

Der Wert 11 des EpubUsageType kommt aus der Codeliste 145, Usage Type 11, „Text and data mining“.

Der Wert EpubUsageStatus kommt aus der Codeliste 146, Usage Status 03, „prohibited“.

Wenn man ausdrücklich das „Text and data mining“ erlauben möchte, kann als Wert für EpubUsageStatus der Wert 01, „Permitted unlimited“ genutzt werden.

Zusätzlich kann eine „Lizenz“ mitgegeben werden, die das Recht für TDM genauer regelt oder eine Lizenz dafür anbietet:

Abschnitt	ONIX-Tags	
	<EpubLicense>	
P.3.20a	<EpubLicenseName>MyPublisher TDM License Policy</EpubLicenseName>	
	<EpubLicenseExpression>	
P.3.20b	<EpubLicenseExpressionType>20</EpubLicenseExpressionType>	20 = ODRL
P.3.20d	<EpubLicenseExpressionLink> https://www.mypublisher.com/.policies/contentpolicy.json </EpubLicenseExpressionLink>	
	</EpubLicenseExpression>	
	</EpubLicense>	

Der Wert 20 ist NEU in Liste 218 in Version 65 der ONIX Codelisten!

Alle Codewerte der Liste 218 – (License expression type)				
Code wert	Bezeichnung	Beschreibung	Version	Update
01	Human readable	Document (eg Word file, PDF or web page) Intended for the lay reader	24	24
02	Professional readable	Document (eg Word file, PDF or web page) Intended for the legal specialist reader	24	24
10	ONIX-PL		24	24
20	ODRL	Open Digital Rights Language (ODRL) in JSON-LD format. Used for example to express TDM licenses using the W3C TDM Reservation Protocol	65	65

Die Möglichkeiten der Lizenzdatei selber wird in Kapitel [1.5](#) erklärt.

1.2.2 Schutz eines Marketingtextes

Wenn ein im ONIX mitgelieferter Text nicht zu KI-Training genutzt werden soll, lässt sich dies folgendermaßen beschreiben:

Abschnitt	ONIX-Tags	
	<TextContent>	
P.14.1a	<TextTyp>03</TextType>	03 = Description
P.14.2	<ContentAudience>03</ContentAudience>	03 = End-customers
P 14.3	<Text textformat="05"><p>Willi Kraus ist der beste Ermittler in Berlin. Als Jude jedoch wird er von seinen Vorgesetzten schikaniert. Als in Berlin immer mehr Kinder verschwinden und an dunklen Orten seltsame Knochen auftauchen, beginnt Kraus zu ermitteln. Buchstäblich im Untergrund der Stadt findet er eine heiße Spur. Dann aber entzieht man ihm den Fall und protegirt einen anderen Polizisten, der sich als Anhänger einer neuen, angeblich patriotischen Partei erweist. Für Kraus wird die Luft im Präsidium immer dünner. Juden gelten plötzlich wieder als Vaterlandsverräter. Doch dann wird der Mordfall immer monströser – und seinen Vorgesetzten bleibt nichts anderes übrig, als Kraus zurückzuholen ...</p></Text>	
	<EpubUsageConstraint>	
P.14.6b	<EpubUsageType>11</EpubUsageType>	11 = Text and data mining
P 14.6c	<EpubUsageStatus>03</EpubUsageStatus>	03 = Prohibited
	<EpubUsageConstraint>	
	</TextContent>	

Auch hier ist eine Lizenz optional (siehe Kapitel [1.5](#)).

1.2.3 Schutz eines über einen Link zur Verfügung gestellten Inhalts

Auch der Inhalt, der über einen Link im ONIX zur Verfügung gestellt wird (Cover, Leseprobe, Verlagslogo, Inhaltsverzeichnis, ...), kann geschützt werden:

Abschnitt	ONIX-Tags	
	<SupportingResource>	
P.16.1a	<ResourceContentType>01</ResourceContentType>	01 = Front cover
P.16.2	<ContentAudience>00</ContentAudience>	00 = Unrestricted
P.16.3	<ResourceMode>03</ResourceMode>	03 = Image
	<ResourceVersion>	
P.16.7	<ResourceForm>02</ResourceForm>	02 = Downloadable file
P.16.11	<ResourceLink>https://s3.eu-central-1.amazonaws.com/public/9999991204707.jpg</ResourceLink>	
	<EpubUsageConstraint>	
P.16.11a	<EpubUsageType>11</EpubUsageType>	11 = Text and data mining
P.16.11b	<EpubUsageStatus>03</EpubUsageStatus>	03 = Prohibited
	<EpubUsageConstraint>	
	<EpubLicense>	
P.16.11e	<EpubLicenseName>MyPublisher TDM License Policy</EpubLicenseName>	
	<EpubLicenseExpression>	
P.16.11f	<EpubLicenseExpressionType>20</EpubLicenseExpressionType>	20 = ODRL
P.16.11h	<EpubLicenseExpressionLink>https://www.mypublisher.com/.policies/mypolicy.json</EpubLicenseExpressionLink>	
	</EpubLicenseExpression>	
	</EpubLicense>	
	<ContentDate>	
P.16.12	<ContentDateRole>17</ContentDateRole>	17 = Last updated
P.16.14	<Date dateformat="00">20220909</Date>	
	</ContentDate>	
	</ResourceVersion>	
	</SupportingResource>	

1.2.4. Schutz eines Covers

Abschnitt	ONIX-Tags	
	<SupportingResource>	
P.16.1a	<ResourceContentType>01</ResourceContentType>	01 = Front cover
P.16.2	<ContentAudience>00</ContentAudience>	00 = Unrestricted
P.16.3	<ResourceMode>03</ResourceMode>	03 = Image
	<ResourceVersion>	
P.16.7	<ResourceForm>02</ResourceForm>	02 = Downloadable file
P.16.11	<ResourceLink>https://cover/public/9783841204707.jpg</ResourceLink>	
	<EpubUsageConstraint>	
P.16.11a	<EpubUsageType>11</EpubUsageType>	11 = Text and data mining
P.16.11b	<EpubUsageStatus>03</EpubUsageStatus>	03 = Prohibited
	<EpubUsageConstraint>	
	<EpubLicense>	
P.16.11e	<EpubLicenseName>AI TDM Conditions</EpubLicenseName>	
	<EpubLicenseExpression>	
P.16.11f	<EpubLicenseExpressionType>20</EpubLicenseExpressionType>	20 = ODRL
P.16.11h	<EpubLicenseExpressionLink> https://Musterverlag.de/richtlinie/richtlinie.json </EpubLicenseExpressionLink>	
	</EpubLicenseExpression>	
	</EpubLicense>	
	</ResourceVersion>	
	</SupportingResource>	

1.3 Schutz außerhalb von ONIX

Auch wenn dieses Best Practice seinen Schwerpunkt auf die Kennzeichen IM ONIX hat, wollen wir die Möglichkeiten des OPT-out darüber hinaus kurz darstellen.

1.3.1 Allgemeiner Schutz im Impressum

Idealerweise befindet sich – wie oben schon aufgeführt – direkt im Content ein Hinweis. Dieser ist zwar weniger maschinell auslesbar, weist aber doch unmittelbar für Lesende (und damit auch KI-Systeme) darauf hin, ob ein Text durch KIs verarbeitet werden darf oder nicht. Er ist insbesondere dann relevant, wenn ein Werk in Papierform digitalisiert werden soll, um dann für TDM verwendet zu werden.

Folgende Texte kommen bereits im Impressum von Verlagen zum Einsatz:

Beispieltext A:

Der Verlag behält sich das Text- und Data-Mining nach § 44b UrhG vor, was hiermit Dritten ohne Zustimmung des Verlages untersagt ist.

Beispieltext B:

Dieser Inhalt ist urheberrechtlich geschützt und darf nicht zur Entwicklung, zum Training oder zur Anreicherung von KI-Systemen, insbesondere generativen KI-Systemen, verwendet werden. Die Nutzung der Inhalte für Text- und Data-Mining ist untersagt.

Empfehlung RA Börsenverein, siehe [„Künstliche Intelligenz aus urheberrechtlicher Sicht - Börsenverein“](#):

Die automatisierte Analyse des Werkes, um daraus Informationen insbesondere über Muster, Trends und Korrelationen gemäß § 44b UrhG („Text und Data Mining“) zu gewinnen, ist untersagt.

1.3.2 Schutz am digitalen Produkt

Der Schutz am oder im digitalen Produkt ist deshalb besonders sinnvoll, weil nicht immer Metadaten zum Produkt vorliegen. Im Folgenden werden einzelne Möglichkeiten dargestellt, wo bzw. ob direkt am digitalen Produkt Kennzeichen für ein TDM-Opt-out hinterlegt werden können. Nicht alle dieser Möglichkeiten sind bisher standardisiert.

1.3.2.1 EPUB

Die Arbeitsgruppen W3C Publishing Maintenance und W3C TDMRep haben für EPUB2 und EPUB3 eine Spezifikation für ein Opt-out herausgegeben.

Die Kennzeichen werden im OPF (content.opf) des EPUBs eingebettet.

Beispielhaft steht dann dies in der OPF-Datei eines EPUB3:

```
<package version="3.0" unique-identifier="BookId" prefix="tdm: https://boersenverein.de/tdm"
xmlns="http://www.idpf.org/2007/opf">
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/" xmlns:opf="http://www.idpf.org/2007/opf">
    <dc:identifier id="BookId">urn:uuid:3d5ea832-4a80-40e9-acd1-8cf57031408a</dc:identifier>
```

```
<dc:language>de-DE</dc:language>
<dc:title>TDM im EPUB</dc:title>
<meta property="tdm:reservation">1</meta>
<meta property="tdm:policy">https://www.musterverlag.de/policies/tdm-policy.json</meta>
</metadata>
</package>
```

Auch hier ist die Policy optional, für das Opt-out ist die Zeile
`<meta property="tdm:reservation">1</meta>`
relevant.

Beispiel der Implementierung für EPUB2:

```
<package version="2.0" ...>
  <metadata ...>
    <dc:title>Document title</dc:title>
    <meta name="tdm:reservation" content="1" />
    <meta name="tdm:policy" content="https://provider.com/policies/policy.json" />
  </metadata>
</package>
```

1.3.2.2 Webseiten und HTML-Darstellungen

Inhalte auf Webseiten und sonstige HTML-Darstellungen (zum Beispiel Web-Applikationen mit Zugriff und Darstellung von Datenbankinhalten) lassen sich mit einem TDM-Opt-out kennzeichnen – und sollen damit die Verarbeitung zum Trainieren einer KI unterbinden.

Das Opt-out für KIs wird in einer separaten Datei `tdmrep.json` im Repository des Webserver abgelegt.

Beispiel für ein Opt-out für alle Dateien:

```
[
  {
    "location": "/",
    "tdm-reservation": 1
  }
]
```

In dem folgenden Beispiel hostet ein Webserver drei Gruppen von Dateien. Der Rechteinhaber der ersten Dateigruppe (PDF-Dokumente) möchte zum Ausdruck bringen, dass die Rechte an diesen Dateien reserviert sind, ohne dass eine TDM-Lizenz erworben werden kann. Der Rechteinhaber der zweiten Gruppe von Dateien (HTML-Seiten) möchte zum Ausdruck bringen, dass die TDM-Rechte mit einer TDM-Policy reserviert sind. TDM-Rechte sind nicht für alle JPEG-Bilder in der dritten Gruppe reserviert.

```
[
  { "location": "/directory-a/",
    "tdm-reservation": 1
  },
  { "location": "/directory-b/html/",
    "tdm-reservation": 1,
  },
  { "location": "/directory-c/jpeg/"
  }
]
```

```
"tdm-policy": "https://provider.com/policies/policy.json"  
},  
{"location": "/directory-b/images/*.jpg",  
 "tdm-reservation": 0  
}  
]
```

Eine ausführliche Beschreibung des W3C zur Umsetzung befinden sich unter <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/>.

1.3.2.3 PDF

Diese Technik bietet eine Möglichkeit, die Wahl des Rechteinhabers in eine PDF-Datei einzubetten.

tdm:reservation und das optionale tdm:policy sind zwei Elemente, die dem XMP-Metadatenabschnitt der PDF-Datei hinzugefügt werden. Sie verwenden beide das Präfix tdm:namespace, das eine Kurzform für die URL <http://www.w3.org/ns/tdmrep/> ist.

Diese Eigenschaften umfassen die TDM-Rechte für jede Seite des Dokuments.

Im folgenden Beispiel signalisiert der Rechteinhaber, dass er sich TDM-Rechte für das PDF-Dokument vorbehält, aber eine TDM-Lizenz erworben werden kann.

```
<rdf:RDF xmlns:rdf=http://www.w3.org/1999/02/22-rdf-syntax-ns#> <rdf:Description rdf:about="" ...  
xmlns:tdm="http://www.w3.org/ns/tdmrep/"> <tdm:reservation>1</tdm:reservation>  
<tdm:policy>https://publisher.com/policies/policy.json</tdm:policy> </rdf:Description> </rdf:RDF>
```

(Quelle: <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240510/#sec-pdf>)

1.3.2.4 Cover (JPG)

Das Cover ist in der Regel urheberrechtlich geschützt und sollte separat zur Verwendung durch KI-Trainingsmodelle freigegeben oder gesperrt werden.

Es gibt ein Kennzeichen für ein KI-Opt-out:

<https://ns.useplus.org/ldf/vocab/DMI-PROHIBITED-AIMLTRAINING> (Prohibited for AI/ML training)

Das IPTC fordert alle Anbieter von Foto-Metadaten-Software auf, die neuen Eigenschaften so schnell wie möglich in ihre Tools zu integrieren, um den Anforderungen der Fotoindustrie gerecht zu werden.

ExifTool (<https://exiftool.org/>), das Befehlszeilentool für den Zugriff auf und die Bearbeitung von Metadaten in Bilddateien, unterstützt die neuen Eigenschaften bereits. Die Unterstützung wurde in der Version 12.67 von ExifTool hinzugefügt, die auf exiftool.org zum Download zur Verfügung steht.

1.4 Nutzungsbedingungen

In allen Bereichen kann eine Erlaubnis der Nutzung des Inhalts ausgesprochen werden. Diese Nutzung kann mit Bedingungen verknüpft werden, die in einem Lizenzhinweis formuliert werden:

Abschnitt	ONIX-Tags	
	<EpubLicense>	
	<EpubLicenseName>MyPublisher Asset TDM Policy</EpubLicenseName>	
	<EpubLicenseExpression>	
	<EpubLicenseExpressionType>20</EpubLicenseExpressionType>	20 = ODRL
	<EpubLicenseExpressionLink> https://www.mypublisher.com/policies/assetpolicy.json</EpubLicenseExpressionLink>	
	</EpubLicenseExpression>	
	</EpubLicense>	

Die Lizenzdatei im .json-Format hat (beispielhaft) diesen Aufbau:

```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://musterverlage.de/policies/1",
  "assigner": {
    "uid": "https://musterverlag.de",
    "vcard:fn": "Musterverlag",
    "vcard:hasEmail": "mailto:contact@musterverlag.de",
    "vcard:hasTelephone": "tel:+4955555555",
    "vcard:hasURL": "https://musterverlag.de/tdm/licensing.html"
  },
  "permission": [{
    "action": "tdm:mime",
    "duty": [{
      "action": "obtainConsent"
    }]
  }]
}
```

Zusammengefasst definieren diese gewählten Ausprägungen ein Angebot, bei dem Text- und Data-Mining erlaubt ist, vorausgesetzt, dass vorher eine Einwilligung eingeholt wird.

Richtlinien werden verwendet, um zulässige und verbotene Aktionen in Bezug auf einen bestimmten Vermögenswert sowie die von den Beteiligten zu erfüllenden Verpflichtungen darzustellen. Darüber hinaus können Richtlinien durch Einschränkungen (z. B. zeitliche oder räumliche Einschränkungen) begrenzt sein und es können Pflichten (z. B. Zahlungen) für Berechtigungen auferlegt werden.

Die Lizenzdatei orientiert sich technisch am Open Digital Rights Language (ODRL, siehe auch <https://www.w3.org/TR/odrl-model/>). Dies ist eine Policy-Expression-Sprache, die ein flexibles und interoperables Informationsmodell, Vokabular und Kodierungsmechanismen zur Darstellung von Aussagen zur Nutzung von Inhalten und Diensten bietet. Das ODRL-Informationsmodell beschreibt die zugrunde liegenden Konzepte, Entitäten und Beziehungen, die die Grundlage für die Semantik der ODRL-Richtlinien bilden.

Beschreibung der Ausprägungen

Das oben aufgeführte JSON-Dokument beschreibt ein Angebot (Offer) im Zusammenhang mit dem Text- und Data-Mining (TDM), unter Verwendung von ODRL (Open Digital Rights Language) und TDM Rep (Text and Data Mining Representation), siehe auch <https://www.w3.org/community/reports/tdmrep/CG-FINAL-tdmrep-20240202/#sec-odrl-profile-context>. Unter <http://www.w3.org/ns/odrl.jsonld> ist beschrieben, welchen Aufbau und welche Ausprägungen die Felder der JSON-Struktur haben (können). Es folgt eine Beschreibung der Felder, damit es leichter fällt, z.B. als Verlag die JSON-Datei zu erstellen.

Feld	Ausprägung	Beschreibung	Empfehlung
context	http://www.w3.org/ns/odrl.jsonld , http://www.w3.org/ns/tdmrep.jsonld	Definiert den Kontext für die JSON-LD-Daten, um deren Bedeutung zu präzisieren. In diesem Fall wird auf ODRL und TDM Rep verwiesen.	Nicht verändern
type	offer	Gibt den Typ des Dokuments an, hier ein "Offer", d.h. Angebot der Nutzung der Richtlinie Neben "Offer" gibt es noch andere Typen wie "Agreement" (Vereinbarung) oder "Policy" (Richtlinie), abhängig davon, was genau beschrieben werden soll.	Nicht verändern
profile	http://www.w3.org/ns/tdmrep	Zeigt an, dass das Dokument dem TDM Rep-Profil entspricht.	Nicht verändern
uid	https://musterverlag.de/policies/1	Eine eindeutige Kennung (Uniform Resource Identifier) für das Angebot.	Individueller Pfad
assigner	Hier werden verschiedene Kontaktinformationen des Anbieters angegeben. Diese Einträge ermöglichen es, den Anbieter des Angebots klar zu identifizieren und zu kontaktieren. Sie sind besonders nützlich, um standardisierte Kontaktinformationen bereitzustellen, die maschinenlesbar und leicht in andere Systeme integriert werden können.		
uid	https://musterverlag.de	Eine eindeutige Kennung des Anbieters	individueller Eintrag

vcard:fn	Musterverlag	Der vollständige Name des Anbieters,	individueller Eintrag
vcard:hasEmail"	mailto:contact@musterverlag.de	Die E-Mail-Adresse des Anbieters	individueller Eintrag
vcard:hasTelephone	tel:+4955555555	Die Telefonnummer des Anbieters	individueller Eintrag
vcard:hasURL	https://musterverlag.de/tdm/licensing.html	Die URL zu einer Webseite des Anbieters	individueller Eintrag
permission	Definiert die erlaubten Aktionen und damit verbundenen Pflichten		
tdm	mine Andere Ausprägungen: <ul style="list-style-type: none"> • tdm:analyze: Analysieren von Daten, um Muster, Trends oder Erkenntnisse zu extrahieren. • tdm:extract: Extrahieren spezifischer Datenpunkte oder Informationen aus größeren Datensätzen. • tdm:transform: Umwandeln von Daten in ein anderes Format oder eine andere Struktur. • tdm:store: Speichern von Daten für zukünftige Analyse Zwecke. • tdm:share: Teilen von analysierten Daten oder Ergebnissen mit anderen Parteien. • tdm:use: Allgemeine Nutzung der Daten für verschiedene Zwecke im Rahmen des TDM. 	Das Text- und Data-Mining ist erlaubt	Für Aufweichung mine beibehalten
duty	Verpflichtend zu erfüllende Voraussetzung folgt		
action	obtainConsent Andere (auch zusätzliche) Ausprägungen: <ul style="list-style-type: none"> • payFee: Nutzung gegen eine Gebühr • provideAttribution: Eine Quellenangabe aufführen. • deleteAfterUse: Erzeugte Daten nach der Nutzung löschen 	Die Pflicht „duty“, die erfüllt werden muss (Einwilligung einholen), um die unter tdm aufgeführte Verwendung zu nutzen.	Nicht verändern

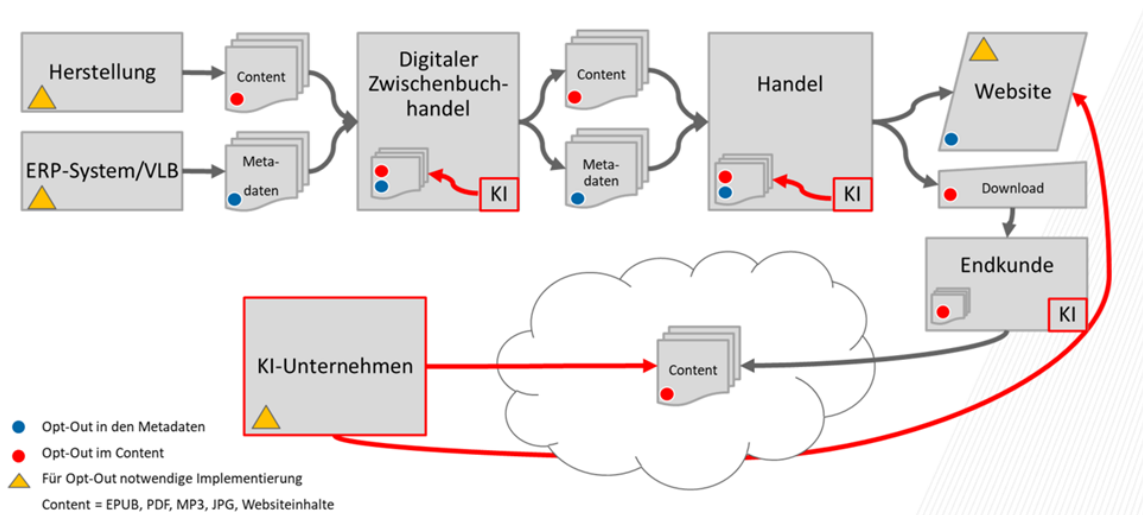
Es folgt ein weiteres Beispiel einer JSON-Datei, die neben der verpflichtenden Einwilligung auch eine Gebühr und die Kennzeichnung der Quelle fordert.

```
{
  "@context": [
    "http://www.w3.org/ns/odrl.jsonld",
    "http://www.w3.org/ns/tdmrep.jsonld"
  ],
  "@type": "Offer",
  "profile": "http://www.w3.org/ns/tdmrep",
  "uid": "https://musterverlage.de/policies/1",
  "assigner": {
    "uid": "https://musterverlag.de",
    "vcard:fn": "Musterverlag",
    "vcard:hasEmail": "mailto:contact@musterverlag.de",
    "vcard:hasTelephone": "tel:+4955555555",
    "vcard:hasURL": "https://musterverlag.de/tdm/licensing.html"
  },
  "permission": [{
    "action": "tdm:mime",
    "duty": [{
      "action": "obtainConsent"
    }],
    {
      "action": "payFee",
      "amount": "10.00",
      "currency": "EUR"
    },
    {
      "action": "provideAttribution",
      "attribution": "Data provided by Musterverlag"
    }
  ]
}]
}
```

1.5 Veranschaulichung der KI-Opt-out-Implementierungen

Technische Umsetzung des AI-Opt-out

Gesamtüberblick



Dieses Schaubild zeigt, an welchen Stellen Opt-out-Implementierungen – jeweils in den Metadaten und direkt im Produkt) erforderlich sind. Dabei gibt es die Implementierung zum **Einbetten** des Kennzeichens und zum anderen die Implementierung des **Auslesens**.

1.6 Fazit Nutzungsvorbehalt

Es ist zu erwarten, dass sich weitere und auch granulare Möglichkeiten zur Anbringung von Nutzungsvorbehalten im Rahmen der Kennzeichnung von Content entwickeln. Auch wenn sich aktuell die Kennzeichnung von Content, der nicht verwendet werden darf (Opt-Out) noch schwierig gestaltet, ermöglicht die ONIX-Kennzeichnung doch bereits jetzt, Inhalte gezielt für KI-Anwendungen freizugeben oder deren Nutzung auszuschließen. So können Content-Ersteller:innen und Verlage proaktiv Urheberrechte im Zeitalter der künstlichen Intelligenz wahren.

2. Kennzeichnung von KI-Inhalten

Künstliche Intelligenz – und insbesondere eine Klasse von generativen KI-Systemen, die auf großen Sprachmodellen basieren, mit denen plausible Texte für eine Vielzahl von Verwendungszwecken erstellt werden können – ist Gegenstand vieler Diskussionen in der Verlagsbranche. Solche Modelle können Texte zusammenfassen, um Inhaltsangaben zu erstellen oder Klassifizierungen vorzuschlagen und sie können zum Schreiben von Marketingmaterial, Artikeln oder ganzen Büchern verwendet werden. Verwandte KI-Modelle können Illustrationen erstellen, Hörbücher vorlesen oder Übersetzungen vornehmen.

Dies führt zu der Sorge, dass eine Masse an KI-generierten Inhalten die von Menschen erstellten Inhalten zu überschwemmen drohen.

Eine Reaktion darauf ist der Verzicht auf den Einsatz von KI oder auf den Handel mit KI-erzeugten Produkten. Eine realitätsnahe Option ist es, gegenüber Handelspartnern und Lesern transparent zu machen, wenn KI eingesetzt wurde. Einige Händler schränken KI-generierte Inhalte ein oder verbieten diese auf ihren Plattformen. Eine Möglichkeit ist es, diejenigen Produkte zu kennzeichnen, bei deren Erstellung generative KI-Techniken verwendet wurden. Es geht hier nicht um die Vor- und Nachteile des Einsatzes von KI in normalen Geschäftsprozessen oder gar bei der Erstellung von Metadaten. Stattdessen möchten wir jenen Verlagen eine Orientierung an die Hand geben, die KI-generierte Inhalte als solche kennzeichnen möchten. Im Fokus steht daher die Frage, wie der Einsatz von KI bei der Erstellung von Produktinhalten in ONIX-Produktdatensätzen ausgedrückt werden kann.

Eine weitere Möglichkeit könnte in der Nutzung einer negativen Kennzeichnung liegen, eine Art "No AI-Label". Über diese Option könnten Content-Ersteller:innen ausdrücken, dass keine generative KI bei der Erstellung der Produkte eingesetzt wurde (besonders wertvoll z.B. bei Fotobänden). Die Problematik hierin liegt in der nicht definierbaren Grenze des Schöpfungsgrads, da Künstliche Intelligenz in diversen Ausprägungen bereits bei der Erstellung und Vermarktung von Produkten Verwendung findet. Wie die Definition der Beteiligung von KI aussehen könnte, um ein "No AI-Label" zu rechtfertigen, ist bisher aber völlig offen.

2.1 KI als Contributor / Urheber*in

Die Beteiligung von KI an der Erstellung von Inhalten kenntlich zu machen, lässt sich äußerst einfach durch die Verwendung von <UnnamedPersons> erreichen.

2.1.1 Covergestaltung

In diesem Beispiel wurde die Covergestaltung eines Buches (digital oder physisch) von einem nicht benannten KI-Modell erstellt. Die ersten beiden Urheber*innen sind beispielsweise Co-Autor*innen und mit SequenceNumber 3 wird nun die ContributorRole Cover Design einsortiert.

Abschnitt	ONIX-Tags	
	<Contributor>	
P.7.1	<SequenceNumber>3</SequenceNumber>	
P 7.2	<ContributorRole>A36</ContributorRole>	A36 = Cover design or artwork by
P 7.20a	<UnnamedPersons>09</UnnamedPersons>	09 = AI (Artificial intelligence)

	<code></Contributor></code>
--	-----------------------------------

Ein wahrscheinlicheres Szenario ist, dass (menschliche) Designer*innen KI-Unterstützung in Anspruch nehmen.:

Abschnitt	ONIX-Tags	
	<code><Contributor></code>	
P.7.1	<code><SequenceNumber>3</SequenceNumber></code>	
P 7.2	<code><ContributorRole>A36</ContributorRole></code>	A36 = Cover design or artwork by
P 7.9	<code><PersonName>Malcolm Smyth</PersonName></code>	
	<code></Contributor></code>	
	<code><Contributor></code>	
P.7.1	<code><SequenceNumber>4</SequenceNumber></code>	
P 7.2	<code><ContributorRole>Z01</ContributorRole></code>	Z01 = Assisted by
P 7.20a	<code><UnnamedPersons>09</UnnamedPersons></code>	09 = AI (Artificial intelligence)
	<code></Contributor></code>	

Zu beachten ist, dass bei der Verwendung von "assisted by" die Sequenznummer und die Rolle der einzelnen Urheber*innen entscheidend ist - im obigen Beispiel gibt es keinen Hinweis auf KI-Unterstützung für die Co-Autor*innen mit den Sequenznummern 1 und 2. Eine ähnliche Vorgehensweise ließe sich für (menschliche) Illustrator*innen von Innenillustrationen anwenden, die KI-generiertes Material als Grundlage für diese Illustrationen verwenden, oder für jede andere Rolle - sogar für menschliche Autor*innen, die KI als Unterstützung bei der Erstellung eines Textes nutzen.

Um einen anderen Schöpfungsgrad auszudrücken, lässt sich die die Sequenznummer und die Rolle der einzelnen Urheber*innen auch umdrehen. In diesem Fall wird ausgedrückt, dass das Produkt der Künstlichen Intelligenz von (menschlichen) Designer*innen nachbearbeitet wurde:

Abschnitt	ONIX-Tags	
	<code><Contributor></code>	
P.7.1	<code><SequenceNumber>3</SequenceNumber></code>	
P 7.2	<code><ContributorRole>A36</ContributorRole></code>	A36 = Cover design or artwork by
P 7.9	<code><UnnamedPersons>09</UnnamedPersons></code>	09 = AI (Artificial intelligence)
	<code></Contributor></code>	
	<code><Contributor></code>	
P.7.1	<code><SequenceNumber>4</SequenceNumber></code>	
P 7.2	<code><ContributorRole>Z01</ContributorRole></code>	Z01 = Assisted by
P 7.20a	<code><PersonName>Malcolm Smyth</PersonName></code>	
	<code></Contributor></code>	

2.1.2. Übersetzung

Durch die Verwendung von <UnnamedPersons> lässt sich ebenfalls die Beteiligung von KI an der Übersetzung von Inhalten kenntlich machen. In diesem Beispiel wurde die Übersetzung von einer menschlichen Übersetzerin mithilfe ("assisted by") eines nicht benannten KI-Modells erstellt.

Abschnitt	ONIX-Tags	
	<Contributor>	
P.7.1	<SequenceNumber>5</SequenceNumber>	
P 7.2	<ContributorRole>B06</ContributorRole>	B06 = Translated by
P 7.9	<PersonName>Zoe Miller</PersonName>	
	</Contributor>	
	<Contributor>	
P.7.1	<SequenceNumber>6</SequenceNumber>	
P 7.2	<ContributorRole>Z01</ContributorRole>	Z01 = Assisted by
P 7.20a	<UnnamedPersons>09</UnnamedPersons>	09 = AI (Artificial intelligence)
	</Contributor>	

2.2 KI-basierte Sprachausgabe in Hörbüchern

In den letzten zehn Jahren wurde die Stimm synthese revolutioniert, indem die Synthese auf der Grundlage von realen Klangbeispielen und der Modellierung von Vokaltrakten durch maschinelles Lernen und Training von neuronalen Netzen ersetzt wurde.

2.2.1 Personalisierte KI-Stimme

Die Verwendung einer KI-generierten Stimme in einem Hörbuch, die auf **einer** menschlichen Stimme basiert, kann folgendermaßen dargestellt werden:

Abschnitt	ONIX-Tags	
	<Contributor>	
P.7.1	<SequenceNumber>2</SequenceNumber>	
P 7.2	<ContributorRole>E07</ContributorRole>	E07 = Read by
P 7.20a	<UnnamedPersons>08</UnnamedPersons>	08 = Synthesised voice – based on real voice actor
	<AlternativeName>	
P 7.21	<NameType>04</NameType>	04 = 'Real' name
P 7.25	<PersonName>Amina Ohanna</PersonName>	
	</AlternativeName>	
	</Contributor>	

Der Code [08](#) bezeichnet eine synthetische oder KI-basierte Stimme, die auf der Stimme einer bestimmten menschlichen Person basiert (trainiert wurde). Das Composite AlternativeName kann dann verwendet werden, um den Namen der sprechenden Person anzugeben.

2.2.2 Nicht personalisierte KI-Stimme

Im Gegensatz dazu sollte für eine nicht personalisierte KI-basierte Stimme, die anhand von Klangbeispielen einer großen Anzahl von Sprecher*innen trainiert wurde, einer der Codes [05-07](#) für unbenannte Personen verwendet werden:

Abschnitt	ONIX-Tags	
	<Contributor>	
P.7.1	<SequenceNumber>2</SequenceNumber>	
P 7.2	<ContributorRole>E07</ContributorRole>	E07 = Read by
P 7.20a	<UnnamedPersons>07</UnnamedPersons>	07 = Synthesised voice – unspecified
	<AlternativeName>	
P 7.21	<NameType>07</NameType>	07 = Fictional character name
P 7.25	<PersonName>Brian Lee Jr.</PersonName>	
	</AlternativeName>	
	</Contributor>	

Alle Codewerte der Liste 19 – (Unnamed person(s))				
Codewert	Bezeichnung	Beschreibung	Version	Update
01	Unknown		0	0
02	Anonymous	Note that Anonymous can be interpreted as singular or plural. A real name can be provided using <AlternativeName> where it is generally known.	0	49
03	Et.al.	And others. Use when some but not all contributors are listed.	0	0
04	Various	When there are multiple contributors, and none are listed individually. Use for example when the product is a pack of books by different authors.	2	49
05	Synthesised voice - male	Use with Contributor role code E07 'read by', eg for audio books with digital narration	8	49
06	Synthesised voice - female	Use with Contributor role code E07 'read by', eg for audio books with digital narration	8	49
07	Synthesised voice – unspecified	Use with Contributor role code E07 'read by', eg for audio books with digital narration	8	49
08	Synthesised voice – based on a real voice actor	Use with Contributor role code E07 'read by', eg for audio books with digital narration, and provide name of voice author in <AlternativeName>. Only for use in ONIX 3.0 or later	49	49
09	AI (Artificial Intelligence)	Use when the creator (of text, of images etc) is a generativ AI model or technique. Note, can also be combined with the role 'assisted by'. Only for use in ONIX 3.0 or later	62	62

2.3 Fazit Kennzeichnung

Immer öfter stellen Kund*innen die Frage, ob Inhalte KI-generiert sind oder nicht. Auch Teile des Marktes, wie Amazon, fordern die Kennzeichnung von KI-generierten Inhalten. Insbesondere KI-generierte Ratgeber, wie ein KI-Buch für Pilzsammler, können fatale Auswirkungen haben. Das vorliegende Best Practice zeigt, dass eine KI-Kennzeichnung für jene Verlage, die KI generierte Inhalte als solche kennzeichnen wollen, in ONIX bereits möglich ist. Eine solche Kennzeichnung bringt aber auch Herausforderungen mit sich, weil der Grad des KI-Beitrags nicht klar definiert ist. Hier besteht noch Diskussionsbedarf.

Wir als Peergroup KI in ONIX empfehlen, dass Inhalte, die im Wesentlichen oder gar ausschließlich ihren Ursprung in einer KI haben, kenntlich gemacht werden. Es geht nicht darum, Inhalte zu kennzeichnen, die von einem Menschen erstellt wurden und von einer KI angepasst oder überarbeitet wurden. Auch Inhalte, die von einer KI erstellt wurden von Menschen mit mehr als 100 Prompts bearbeitet wurden, müssen nicht zwingend gekennzeichnet werden. Es geht darum, überhaupt erste KI-Kennzeichnungen mitzugeben. Käufer*innen werden die Vorteile der Kennzeichnungen schnell zu interpretieren wissen – und manche wollen gerade auch einen KI-generierten Inhalt, weil dieser durchaus Mehrwerte bieten kann. Auf der anderen Seite wollen Händler nicht mit rein KI-generierten Inhalten überschüttet werden, weil diese z.T. qualitative Mängel ausweisen. Die Umsetzung im Handel ist der Kennzeichnung im ONIX durch die Verlage nachgelagert – und kann auch erst dann sein Potenzial entfalten, wenn Händler mit der Umsetzung gestartet sind.

