

# Mögliche Herausforderungen im Bereich OER-Metadaten

In diesem Dokument wollen wir einen Ausblick wagen, welche Herausforderungen für die kommenden Jahre im Bereich OER-Metadaten möglicherweise anstehen. Die Auslistung wurden durch Sichtungen von Veröffentlichungen im OER-Umfeld, durch Analyse bisheriger Aktivitäten und durch Interviews, die primär im Rahmen der gemeinsamen OER-Metadatengruppe von DINI-AG KIM und Jointly geführt wurden.

Wir erheben bei dieser Zusammenstellung nicht den Anspruch auf Vollständigkeit, sondern wollen mit diesem Artikel viel mehr an die Leser:innen appellieren, diese Liste zu vervollständigen / korrigieren, ganz nach Motto "sharing is caring".

Autoren:

- Frank Thomschke
- Steffen Rörtgen

**Ihr habt weitere Herausforderungen identifiziert?  
Gerne vervollständigen, anmerken, weiterschreiben!**

## Mögliche Herausforderungen im Bereich OER-Metadaten

Herausforderung 1: Kompetenzbasierte Lernprozesse unterstützen

Herausforderung 2: Konsistenz der Metadaten beim OER-Remix sicherstellen

Herausforderung 3: Transparenz und Qualität der OER-Wiederverwendung verbessern

Herausforderung 4: Zielgerichtete Verbreitung der Metadaten innerhalb der Community ermöglichen

Herausforderung 5: Kompatibilität zwischen OER & NoOER Metadaten herstellen

Herausforderung 6: Ermittlung und Verknüpfung von Metadaten mittels KI-Unterstützung optimieren

Herausforderung 7: Verknüpfung der Metadaten mit externen Wissensquellen ausbauen

Herausforderung 8: Transparenz der OER-Produktion verbessern

# Herausforderung 1: Kompetenzbasierte Lernprozesse unterstützen

## *Welchen Mehrwert können maschinenlesbare Kompetenzmodelle erzeugen?*

Die Einführung der Kompetenzorientierung spiegelt die generelle Umstellung unseres Bildungssystems von der Input- zur Output-Orientierung wider.

Es geht nicht mehr um die Beschreibung von Lehrinhalten (**“Input“**), sondern vielmehr um das, was der Lernende am Ende einer Lerneinheit in der Lage ist zu tun (**“Output“**) - sowohl aus **kognitiver Sicht** (theoretische Kenntnisse), als auch aus **funktionaler Sicht** (deren praktische Umsetzung).

Generell unterteilt man die Kompetenzen in:

- **Fachkompetenzen** (Mathematik, Deutsch, etc.) und
- **Fächerübergreifende/überfachliche Kompetenzen** (Sprach, Medien-Demokratiebildung, Gesundheitsförderung, Gewaltprävention etc.).

Pro Bereich werden die Kompetenzen meist in einer Taxonomie hierarchisch eingruppiert, um ein **Kompetenzstrukturmodell** zu bilden.

Um Kompetenzen klassifizierbar (damit mess- und vergleichbar) zu machen, werden **Kompetenzstufenmodelle** entwickelt, die eine formale Beschreibung von Schüler:innenkompetenzen ermöglichen.

Im Schulbereich hat man sich z. B. auf ein 5-stufiges Kompetenzstufenmodell (Beispiel [hier](#) auf S. 6) geeinigt, da dieses zu anderen nationalen und internationalen Modellen vergleichbar ist. Die Kompetenzmodelle werden hier institutionell auf drei Ebenen erarbeitet:

1. Die **nationalen Bildungsstandards** der KMK legen (auf Basis der mittleren Kompetenzstufe 3 = Regelstandard) fest, welche Kompetenzen Schüler bis zu einer bestimmten Jahrgangsstufe erworben haben sollen.
2. Die Aufgabe der **länderspezifischen Rahmenlehrplänen** ist es, diese nationalen Vorgaben schulart- und jahrgangsstufenbezogen zu strukturieren, für die Anforderungssituationen im Unterricht aufzubereiten und länderspezifische Vorgaben zu ergänzen.

3. Zum Schluss werden die mit dem Rahmenlehrplan und die mit dem jeweiligen Schulprogramm verbundenen Vorgaben in ein schlüssiges pädagogisches Handlungskonzept – das **schulinterne Curriculum** – überführt.

Natürlich ist es wünschenswert, auch Bildungsmaterialien auf Basis ihrer **Metadaten diesen Kompetenzmodellen zuzuordnen**. Als Beispiel kann hier der [ISQ Kompetenzbrowser](#) genannt werden, der bei Auswahl einer Kompetenz passendes Aufgabenmaterial aus dem [ISQ VERA-Aufgabenpool](#) anbietet.<sup>1</sup>

Der **Kompetenzbezug** in den Metadaten kann auf zwei Arten realisiert werden:

- der Kompetenzbezug wird **per Referenz** hinterlegt. Dazu ist nur ein eindeutiger Identifier für die jeweilige Kompetenz erforderlich. Alle weiteren Detailinformationen müssen bei Bedarf aus dem Kompetenzmodell entnommen werden.

Der [LOM-Standard](#) sowie auch der [LRMI-Standard](#) sehen dafür entsprechende Beschreibungselemente vor, mit denen ein Referenzbezug zu einem Klassifikationssystem hergestellt werden kann. Beim LRMI-Standard wird dies sogar nochmal untersetzt: Je nach Beschreibungselement stellt die referenzierte Kompetenz eine Voraussetzung bzw. das eigentliche Lernziel für das zugehörigen Bildungsmaterial dar oder präsentiert das Kompetenzniveau, das mit dem zugehörigen Bildungsmaterial abgeprüft werden kann.

- der Kompetenzbezug wird **als Wert** hinterlegt, indem alle nötigen Kompetenzinformationen als Kopie in die Metadaten übernommen werden. Dieser Ansatz wird z. B. beim Applikationsprofil [LOM-CH](#) mit einem dedizierten Beschreibungselement für Curricula verfolgt.

Einerseits stehen alle kompetenzbasierten Informationen ohne zusätzliche Abfrage des Kompetenzmodelles sofort zur Verfügung, andererseits müssen alle Metadaten immer dann aktualisiert werden, wenn am Kompetenzmodell Veränderungen vorgenommen werden.

Den entscheidenden Mehrwert erhält man aber erst, wenn nicht nur die einzelnen Kompetenzmodelle für sich spezifiziert werden, sondern **auch ihre Beziehungen untereinander**, um:

- die **Nachhaltigkeit** von Metadaten zu verbessern, indem bei Aktualisierung von Kompetenzmodellen (Lehrplananpassungen) die Bezüge in den Metadaten unverändert bleiben und stattdessen die alten und neuen

---

<sup>1</sup> Eine Anfrage über die Nutzbarkeit des Tools sowie des zugrundeliegenden Schemas läuft.

Kompetenzen der beiden Modelle miteinander verknüpft werden.

- auch **Materialien auch deren Beziehungen** (und nicht nur einzelnen Kompetenzen) **zuzuordnen**. Als Beispiel kann hier die [Curriculum-Plattform](#) von Rheinland-Pfalz genannt werden, die bei Auswahl einer Fachkompetenz weitere überfachliche Kompetenzen inkl. Anmerkungen und Materialien anbietet.
- den **Suchkontext zu erweitern**, indem man verknüpfte Kompetenzen (z. B. anderer Bundesländer oder des nationalen Bildungsstandards) ermitteln und das zugehörige Material bei der Suche mit einfließen lässt.

Dies setzt aber voraus, dass **alle Kompetenzmodelle einheitlich spezifiziert und unabhängig von den Metadaten abrufbar** sind.

Als erfolgreiche Beispiele dafür kann genannt werden:

- Norwegen, die ihre Lehrpläne maschinenlesbar in unterschiedlichen Formaten zentral und mit freier Nutzungslizenz zur Verfügung stellen.<sup>2</sup>
- die europäische Klassifikation [ESCO](#) für Fähigkeiten, Qualifikationen und Berufe (Umfang: 2942 Berufe, 13485 Fähigkeiten und 9457 Qualifikationen aus 10 EU Ländern), die auf Basis des SKOS-Standards definiert wurde.<sup>3</sup>

---

<sup>2</sup> <https://www.udir.no/om-udir/data/kl06-grep>

<sup>3</sup> <https://ec.europa.eu/esco/resources/data/static/model/html/model.xhtml>

# Herausforderung 2: Konsistenz der Metadaten beim OER-Remix sicherstellen

## *Wie könnten wir ein stabiles Remix-Netzwerk aufbauen?*

Zur Definition von "Offenheit" im OER-Umfeld wird häufig auf die 5 Freiheiten nach David Wiley (**Ver**wahren & **Ver**vielfältigen, **Ver**wenden, **Ver**arbeiten, **Ver**mischen, **Ver**breiten) verwiesen. Um diese Freiheiten im Rahmen des OER Workflows voll unterstützen zu können, müssen entsprechende **Grundprinzipien** eingehalten werden, um OER-Materialien auffindbar, zugänglich, interoperabel und wiederverwendbar zu machen.

Eine besondere Herausforderung stellt dabei das **Vermischen (OER-Remix)** dar. Der OER-Produktionsprozess sollte dafür sorgen, dass alle wiederverwendeten/zitierten OER-Materialien in den Metadaten des neuen OER-Materials vermerkt werden. Die Konsistenz dieses so entstehenden "Remix-Netzwerkes" muss nachhaltig sichergestellt werden.

Eine Möglichkeit, um die Konsistenz der Metadaten von veröffentlichten OER Materialien zu fördern, wäre ein **gemeinsames OER-Logbuch** zu etablieren, in dem die URL für den Zugriff auf das Material und die Metadaten zu allen OER Veröffentlichungen (und ggf. deren Widerrufen) **dauerhaft** und **unabhängig von den Datenquellen** hinterlegt werden können. Dadurch wäre sichergestellt, dass jedes veröffentlichte OER Material nachhaltig auffindbar und zitierbar bleibt.

Im Hochschulbereich wird diese Thematik bereits seit längerem unter den Begriffen "Open Data" und "Open Science" diskutiert. Damit Daten diesen Kriterien der Offenheit genügen, wird sich an den FAIR Data Prinzipien (Findable, Accessible, Interoperable, Reusable) orientiert. Diese Prinzipien lassen sich auf das OER-Umfeld übertragen, hier ein Auszug der wichtigsten Prinzipien (inkl. Lösungsansätze):

### **a. Metadaten wird ein global eindeutiger und dauerhaft persistenter Identifier zugewiesen**

Die FAIR Data Prinzipien empfehlen den Einsatz von PID-Systemen, um digitalen Objekten einen weltweit eindeutigen Identifier unabhängig von der Datenquelle zuzuordnen. Dabei können neben der URL für den Content-Zugriff auch die gesamten Metadaten hinterlegt und später abgerufen werden.

Im Internet werden unterschiedliche PID-Systeme bereitgestellt, die sich hinsichtlich Offenheit, Kosten, Aufwand, Verbreitung bzw. Mehrwerte unterscheiden. Laut re3data.org sind das DONA Handle System und der

darauf aufbauende Digital Object Identifier aktuell die verbreitetsten Lösungen im Bereich Forschung und Lehre.

### b. Daten werden mit umfangreichen Metadaten beschrieben

Wichtig für die Transparenz der OER-Produktion (siehe Herausforderung 8) und die nachgelagerte OER-Wiederverwendung (siehe Herausforderung 3) ist auch, von wem das OER-Material erstellt, bearbeitet bzw. veröffentlicht wurde. Bezeichner wie Vor-/Nachname bei Autoren bzw. der Organisationsname einer Redaktion sind dazu aber viel zu ungenau und nicht nachhaltig stabil genug. Deshalb werden auch für Personen und Organisationen stabile Identifier (aber nicht zwingend persistente Identifier) benötigt:

- Für Personen dürfen diese Identifier keine personenbezogene Daten beinhalten bzw. deren Abfrage ohne Einwilligung der Person erlauben.
- Seit 2012 wird von der Non-Profit-Organisation ORCID solche Identifier herausgegeben, um Personen im Wissenschaftssystem weltweit eindeutig identifizieren zu können. Auch Domain of One's Own Services könnten mit ihren personengebundenen Domains einen stabilen Identifier (z.B. als WebID) liefern.
- Für Organisationen könnte man Verweise auf Wikidata, die Gemeinsame Normdatei (GND) bzw. auf die Research Organization Registry verwenden.

### c. Metadaten enthalten qualifizierte Referenzen auf andere Metadaten

Um qualifizierte Relationen zwischen OER-Materialien auszudrücken, wird ein Metadatenstandard mit entsprechenden Beschreibungselementen benötigt, wie z. B. "basierend auf" oder "hat Bestandteil". Beide etablierten Metadaten-Standards im OER-Umfeld (LOM/LRMI) stellen dafür entsprechende Beschreibungselemente in unterschiedlicher Granularität zur Verfügung.

Natürlich müssen im Rahmen der OER-Produktion, indem ein OER-Remix durchgeführt wird, die Metadaten der wiederverwendeten OER-Materialien zur Verfügung stehen (siehe Herausforderung 8), um sie in die Metadaten des neuen OER-Materials mit einfließen zu lassen.

Die Angabe eines persistenten Identifiers würde hierzu vollkommen ausreichen, da die hinterlegten Metadaten bei Bedarf vom PID-System

abgerufen werden können.

### d. Metadaten enthalten eine eindeutige, zugreifbare Angabe einer Nutzungslizenz

Jedes OER-Material muss per Definition mit einer freien Nutzungslizenz versehen werden. Der Remix solcher Materialien, sprich der Mix von freien Nutzungslizenzen, ist leider nicht immer kompatibel zueinander.

Deshalb ist es sehr wichtig, dass einerseits alle Bestandteile als Referenzen (siehe c), andererseits auch die Nutzungslizenzen in den Metadaten hinterlegt werden, um bei der Veröffentlichung die Nutzungslizenzen der einzelnen Bestandteile auslesen und auf deren Basis nur kompatible Nutzungslizenzen für das neue OER Material zur Auswahl anbieten zu können (siehe dazu auch Herausforderung 5).

### e. Metadaten sind/bleiben verfügbar, auch für den Fall, dass die zugehörigen Daten nicht mehr vorhanden sind

Damit das durch Referenzierung (siehe c) entstehende OER-Metadaten-Netzwerk konsistent bleibt, müssen alle Metadaten nachhaltig vorhanden bleiben, auch wenn das zugehörige OER-Material nicht mehr verfügbar ist bzw. dessen Veröffentlichung widerrufen wurde.

Durch Einsatz von persistenten Identifiern können die Metadaten der veröffentlichten OER Materialien separat von den Datenquellen verwaltet und dadurch die geforderte Nachhaltigkeit sichergestellt werden (siehe a).

Natürlich wirft so ein zentraler PID-Service auch kritische Fragen auf, speziell zu:

- **Verfügbarkeit:** Eine Hochverfügbarkeit kann durch PID-Replikation innerhalb eines Konsortiums von Handle/DOI Providern hergestellt werden.
- **Flexibilität:** Jede PID-Instanz kann immer nur ein spezifisches Metadaten-Schema unterstützen. Wenn man sich nicht auf ein gemeinsames OER-Metadaten-Schema einigen kann, dann könnte pro Bildungssektor (Schule, Hochschule etc.) und/oder pro Bundesland eine separate PID-Instanz aufgebaut werden, welcher das gewünschte OER-Metadaten-Schema forciert. Dadurch würde das gemeinsame OER-Logbuch auf mehrere PID-Instanzen mit getrennten Namensräumen aufgeteilt ("data sharding") werden.
- **Offenheit:** Zentrale Dienste bringen immer einen Lock-In-Effekt mit sich. Um dem entgegenzuwirken, wurde z. B. die unabhängigen DONA Foundation als

Non-Profit Organisation gegründet, um die vorgestellten PID-Systeme auf Basis von Mitgliederentscheiden zu regulieren.

- **Abdeckung:** Natürlich sollte jede OER Veröffentlichung im gemeinsamen OER-Logbuch hinterlegt werden, was primär von den Datenquellen vorgenommen werden sollte, von denen ursprünglich die Veröffentlichung ausgegangen ist. Für große proprietäre Plattformen wie YouTube oder Slideshare lässt sich dies aber kaum umsetzen. Hier könnten Dritte die Metadaten von diesen Datenquellen einsammeln und stellvertretend die Eintragung vornehmen.

Alternativ zu einem gemeinsamen OER-Logbuch könnte jede Datenquelle eigene stabile URLs für jede veröffentlichte Version eines OER Materials inkl. Metadaten zur Verfügung stellen. Bei einer neuen Veröffentlichung müsste die jeweilige Datenquelle selbständig die Konsistenz der Metadaten aller wiederverwendeten / zitierten OER Materialien überprüfen und bei einem Widerruf einer Veröffentlichung sicherstellen, dass über die herausgegebene URL weiterhin auf die Metadaten zugegriffen werden kann. Des Weiteren muss nach einem Umzug sichergestellt werden, dass die alte Domain weiterhin auf die Datenquelle zeigt und Änderungen im Pfad müssen intern mittels Weiterleitungen abgefangen werden.

Eine Verbreitung der veröffentlichten Metadaten könnte bei diesem Ansatz durch Aufbau zentraler Suchindizes erfolgen, welche die Metadaten der einzelnen Datenquellen regelmäßig abfragen und die Änderungen über einen event-basiertes API zur Verfügung stellen. Eine kausale Konsistenz kann bei diesem Ansatz aber nicht sichergestellt werden, d.h. die Reihenfolge der generierten Events muss nicht unbedingt mit der wirklichen Reihenfolge der Veröffentlichungen / Widerrufen innerhalb der OER-Infrastruktur übereinstimmen.

Es gibt auch Überlegungen, alle OER-Veröffentlichungen in Zukunft als "smart contracts" in einer Blockchain abzulegen (wodurch die kausale Konsistenz erhalten bleibt) und diesen dezentralen Storage für den gesamten Datenabgleich zwischen allen Services einer dezentralen OER-Infrastruktur zu nutzen.<sup>4</sup>

---

<sup>4</sup> Axel Dürkop: <http://doi.org/10.15480/882.1970>

## Herausforderung 3: Transparenz und Qualität der OER-Wiederverwendung verbessern

*Wie können wir die Wiederverwendung mit mehr Transparenz im OER Umfeld verbessern?*

Wenn Herausforderung 2 umgesetzt wird und damit:

- jedes Material bzw. jede:r Autor:in/Redaktion einen stabilen Identifier besitzt und
- die Wiederverwendung/Zitation anderer Materialien in den Metadaten abgebildet wird,

dann könnten bei der Veröffentlichung eines OER-Materials die **zugehörigen Metadaten ausgewertet** werden, die **Wiederverwendung/Zitation anderer OER-Materialien extrahiert** und damit **Indizes aufgebaut** werden, um eine Rückwärtssuche anzubieten, welche pro Material bzw. pro Autor:in/Redaktion die bestehenden Wiederverwendungen bzw. Zitationen liefern.

Diese generierten Indizes können:

- einerseits für **Recherchen** genutzt werden: Beispielsweise könnte eine Ursprungsautorin abfragen, welches ihrer Materialien von anderen Autor:innen/Redaktionen in ihren Werken wiederverwendet bzw. zitiert wird. Auf Basis dieser Informationen könnten sich Interessengemeinschaften ("Special Interest Groups") im OER-Autor:innenumfeld herausbilden.
- andererseits als **Metrik** für das Material bzw. den/die Autor:in/die Redaktion herangezogen werden (z. B. der Indexumfang = "usage").

Beim Widerruf einer Veröffentlichung müssten die Indizes aktualisiert und die Autor:innen/Redaktionen der Materialien informiert werden, die das widerrufene Material zitiert bzw. wiederverwendet haben, um ihnen die Chance zu geben, zeitnah ihre Materialien anzupassen und damit die Qualität ihrer Materialien wiederherzustellen.

Die Metadaten aller veröffentlichter OER-Materialien könnten dafür aus einem gemeinsamen OER-Logbuch entnommen werden (siehe Herausforderung 2). Alternativ dazu könnte jedes wiederverwendete/zitierte OER-Material auch direkt über diesen Sachverhalt informiert werden, was z. B. mit Hilfe des Webmention-Protokolls oder über Kommunikationskanäle im Rahmen einer Social Media Plattform (siehe Herausforderung 4) realisiert werden kann.

### Herausforderung 4: Zielgerichtete Verbreitung der Metadaten innerhalb der Community ermöglichen

*Wie können wir die OER Community zielgerichtet über Veröffentlichungen informieren?*

Alle kennt das Prinzip von sozialen Netzwerken (Facebook, Twitter, Mastodon & Co.): Man folgt einer Person (weil man ihr vertraut) und diese "Follower" werden gezielt mit den Veröffentlichungen dieser Person versorgt. Man kann diesen "Vertrauenkreis" erweitern, indem man die Personen hinzunimmt, die eine Vertrauensstellung zu den Personen des inneren Vertrauenkreises besitzen. Somit entstehen in sich geschachtelte Vertrauensstellungen ("Circles of Trust").<sup>5</sup>

Eine gezielte Distribution von OER-Metadaten könnte auch auf Basis solcher Vertrauensstellungen erfolgen. Einer OER-Akteurin müsste dazu angeboten werden, einem anderen Akteur "zu folgen". Dabei müssen die OER-Akteur:innen nicht immer die Ursprungsautor:innen der Materialien bzw. nachgelagerte Redaktionsteams sein, sondern es gibt auch OER-Akteur:innen, die Materialsammlungen zu einem speziellen Thema kuratieren.

Es gibt eine Menge etablierter sozialer Netzwerke, die aber meistens die gewünschte Offenheit im OER Umfeld nicht gewährleisten.

ActivityPub ist ein offenes Protokoll für dezentrale soziale Netzwerke, das offiziell als W3C-Standard herausgegeben wurde. Zweck dieser Spezifikation ist es, die Interaktionen auf Social-Network-Plattformen zu standardisieren, also beispielsweise das gegenseitige Folgen, das Liken und das Schicken von Nachrichten. Auf Basis von ActivityPub kann jede Organisation eine eigene ActivityPub-Instanz für seine Nutzer:innen betreiben und sich mit anderen ActivityPub-Instanzen zu einer Föderation zusammenschließen, ohne dass dafür ein zentraler Dienst benötigt wird. Ein prominentes Beispiel für so eine "Open Social Media Plattform" ist der Twitter-Clon Mastodon.

Moodle.Net greift diesen offenen Ansatz auf und will eine "Open Social Media Platform for educators" aufbauen, um Sammlungen von freien Bildungsmaterialien gemeinsam kuratieren zu können. Dazu können die Nutzer sich zu "Communities" zusammenschließen, die "Collections" von freien Bildungsmaterialien herausgeben.

---

<sup>5</sup> Auch die norwegische Plattform NDLA arbeitet an ähnlichen Community-basierten Ansätzen, siehe <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-169937> Kapitel 19.

Zu jeder "Einzelperson/Community/Collection"<sup>6</sup> wird ein entsprechender Kommunikationskanal in der ActivityPub-Föderation aufgemacht.

Zusätzlich kann mit Hilfe des Projektes [SkoHub](#) eine Instanz innerhalb einer ActivityPub-Föderation betrieben werden, die auf Basis von SKOS-Spezifikationen für jeden Deskriptor eines kontrollierten Vokabulars einen eigenen Kommunikationskanal zur Verfügung stellt. Somit wäre es möglich, bei einer Veröffentlichung die Metadaten auszulesen und für jeden extrahierten Deskriptor eine Nachricht auf dem entsprechenden Kommunikationskanal abzusenden. Andererseits könnte auch darüber informiert werden, wenn eine Veröffentlichung widerrufen wird.

Würde man so eine SkoHub-Instanz in der Moodle.Net-Föderation betreiben, könnten sich die Communities bei Interesse an diesen "Metadaten-getriebenen" Kommunikationskanal anmelden und zeitnah über jede neue Veröffentlichung bzw. deren Widerruf informiert werden (siehe [hier](#)). Somit wären die Communities in der Lage, ihre kuratierten Collections immer auf dem aktuellen Stand zu halten.

Wenn jedes Material bzw. jede:r Autor:in/jede Redaktion einen stabile Identifier besitzt (siehe [Herausforderung 2](#)), dann könnte man auf Basis dieser Identifier ebenfalls Kommunikationskanäle für Autor:innen/Redaktionen anbieten und bei OER-Veröffentlichungen mit entsprechenden Nachrichten füllen.

Die Metadaten aller veröffentlichter OER-Materialien müssten dafür aus einem gemeinsamen OER-Logbuch entnommen (siehe [Herausforderung 2](#)) und als [event-getriebenes API](#) zur Verfügung gestellt werden.

---

<sup>6</sup> <https://www.w3.org/TR/activitypub/#actor-objects>

# Herausforderung 5: Kompatibilität zwischen OER & NoOER Metadaten herstellen

*Welche Rahmenbedingungen müssen beachtet werden, wenn OER- und NoOER-Material nebeneinander angeboten werden soll?*

Das Wesen von OER Materialien liegt darin, dass sie legal und kostenfrei vervielfältigt, verwendet, angepasst und verbreitet werden können. Dass dies traditionelle Wertschöpfungsketten in Frage stellt und für die kommerziellen Verlage damit einen Angriff auf ihr bisheriges Geschäftsmodell darstellt, ist sicherlich unbestritten. Ein prominentes Beispiel für diesen Interessenkonflikt ist die norwegische OER-Plattform [NDLA](#), wo nach über 10 Jahren immer noch keine Zusammenarbeit mit kommerziellen Verlagen etabliert werden konnte<sup>7</sup>.

Eine Bildungsinfrastruktur, die zu 100% auf freien Materialien (inkl. Metadaten) basiert, würde einerseits die technologische Umsetzung deutlich vereinfachen, weil z. B. eine Zugriffskontrolle komplett entfallen kann sowie [DRM](#)-Schutzmechanismen in Verbindung mit Abrechnungsverfahren für Bildungsmaterialien nicht erforderlich sind.

Andererseits liegen (noch) nicht genügend freie Bildungsmaterialien vor, um einen Grundstock für eine offene Bildungsinfrastruktur anzubieten. Deshalb sollten Metadaten von freien und kommerziellen Bildungsmaterialien gemeinsam vorgehalten werden, um den Akteuren das volle Spektrum an Bildungsmaterialien anbieten zu können. Voraussetzung dafür ist, dass generell **alle Metadaten einer freien Nutzungslizenz unterliegen**, egal ob von freien oder kommerziellen Anbieter:innen.

Entweder wird ein gemeinsamer Metadaten-Standard definiert, der bei Bedarf für freie und kommerzielle Materialien erweitert werden kann oder es wird ein Mapping zwischen den unterschiedlichen Metadaten-Standards erstellt, welches im Idealfall eine 1:1 Abbildung ermöglicht.

Egal ob gemeinsame oder gemaapte Metadaten, es sollte auf jeden Fall ein Beschreibungselement existieren, mit dem:

- eine **klare Unterscheidung zwischen freien und kommerziellen Materialien** möglich ist.

Nur so können beim Veröffentlichen von OER Materialien entsprechende Warnungen bzw. Sperren angeboten werden, wenn fälschlicherweise

---

<sup>7</sup> <http://nbn-resolving.org/urn:nbn:de:0111-pedocs-169937>

kommerzielle Materialien beim OER Remixing eingeflossen sind.

- ausgedrückt werden kann, ob das Material **lizenzfrei oder mit einer Nutzungsbeschränkung behaftet** ist.

Da "Lizenzfrei nicht gleich freie Lizenz ist"<sup>8</sup>, ist es gerade im OER Umfeld wichtig, die verwendete Nutzungslizenz genau anzugeben. Bei den Creative Commons Lizenzen werden dazu entsprechende URLs zur Verfügung gestellt, die für ein kontrolliertes Vokabular verwendet werden können.

Nur so können beim Veröffentlichen von OER Materialien entsprechende Warnungen bzw. Sperrungen angeboten werden, wenn fälschlicherweise eine nicht-kompatible OER Nutzungslizenz für das neue OER Material ausgewählt wurde.

---

<sup>8</sup> <https://youtu.be/o70I0P6vmeE>

# Herausforderung 6: Ermittlung und Verknüpfung von Metadaten mittels KI-Unterstützung optimieren

*Wo kann sich die OER-Community KI-Unterstützung im Bereich Metadaten vorstellen, welche Projekte gibt es bereits?*

Im Jahr 2019 ist von Jointly ein [Community-Dokument](#) zu dem Thema "Künstliche Intelligenz für Open Educational Resources" gestartet worden. In diesem Rahmen wurden einerseits Ideen gesammelt, wofür sich KI im Bereich OER nutzen ließe und andererseits aktuelle Projekte erfasst, die sich mit dem Thema KI und dessen Einsatzmöglichkeiten für freie Bildungsmaterialien befassen.

Dabei sind fünf Themenbereich identifiziert worden:

- Metadaten
- Inklusion
- LMS u.a. Lernumgebungen
- Content-Generierung
- Chat-Bots und Lizenzierungsassistenten

Im Bereich "Metadaten" sind hierbei vor allem die automatische Generierung von Metadaten (bspw. das Hinzufügen des Themas, Einordnung in einen fachlichen Bereich), das Verknüpfen mit ähnlichen Materialien aus einem Repository und bessere Suchempfehlungen gewünscht worden.

Bei einer Recherche zu bereits bestehenden Projekten, konnten sechs Projekte identifiziert werden, die sich diesem Thema mit Methoden der KI (1x EU, 3x national, 2x Niederlande) widmen:

- [X5GON](#) (EU-Projekt): Die Entwicklung einer KI-getriebenen Plattform soll OER-Inhalte übergreifend für Lernende auffindbar machen. Dabei sollen die Inhalte durch KI kuratiert werden.
  - **Einschätzung:** Momentan werden auf der Website 3-Case-Studies vorgestellt, die sich alle mit dem Thema "Übersetzung von Videocontent" befassen. Außerdem sollen drei Services entwickelt werden, von denen sich "X5oerfeed", vor allem mit der Analyse des Contents zu beschäftigen scheint. Ergebnisse zu diesem Service-Teil scheinen noch nicht vorzuliegen (Anfrage läuft).
- [EduArc](#) (BMBF): Das Projekt EduArc (Laufzeit 10/2018 – 09/2023) beschäftigt sich mit der Erforschung technischer, didaktischer und organisatorischer Gelingensbedingungen einer Infrastruktur, die es ermöglicht, OER

hochschulübergreifend auffindbar zu machen. Ein Teilprojekt (Digitale Bildungsarchitekturen. Offene Lernressourcen in verteilten Lerninfrastrukturen) wird dabei vom DIPF koordiniert und befasst sich unter anderem im AP 7 mit dem Themenfeld "Data Mining".

- **Einschätzung:** Ergebnisse zu dem Arbeitspaket sind noch nicht veröffentlicht.
- SkoHub (hbz): SkoHub ist ein Projekt des Hochschulbibliothekszentrums NRW, um die Auffindbarkeit von Materialien im Web zu verbessern. Dabei wird das Social Networking Protokoll ActivityPub sowie LinkedData Notifications eingesetzt. Mittels Browser-Extension können Webseiten (bspw. OER-Materialien) an die Inbox eines definierten Vokabulars gesendet werden. Interessierte Menschen können diese Inbox bspw. auf ActivityPub basierten sozialen Netzwerken wie Mastodon abonnieren und erhalten so direkt einen Eintrag in ihrer Timeline, wenn ein Material hinzugefügt wurde. In diesem Ansatz zeigen sich Überschneidungen zu MoodleNet. SkoHub plant für jedes Vokabular einen suggest-Endpoint anzubieten, der mit Hilfe des Tools KI-Tools Annif umgesetzt werden soll.
  - **Einschätzung:** Das Tool SkoHub funktioniert bereits. Eine Vorstellung findet sich bspw. auf YouTube. Der KI-Teil befindet sich noch in der Entwicklung und konnte noch nicht getestet werden.
- Schulkontext.de - Die Welt als Schulbuch: Dieses Projekt versucht mittels KI-getriebener Technologien aktuelles Tagesgeschehen mit Rahmenlehrplänen in Verbindung zu bringen und Lehrer:innen bei ihrer Unterrichtsplanung zu unterstützen. Dabei sollen auch OER eingebunden werden. Auch das automatisierte Erstellen von Aufgaben wird angestrebt.
  - **Einschätzung:** Das Projekt ist auf Grund verschiedener Aspekte interessant:
    - Dem Mapping auf die Rahmenlehrpläne (maschinenlesbare Abbildungen der Lehrpläne)
    - Die Einbindung von OER (Crawling und Klassifikation)Eine Demo des Services ist momentan noch nicht verfügbar, soll aber im Laufe des Jahres 2020 vorgestellt werden.
- SURF - (OER)/Education-Suchportal (Niederlande): Ein Projekt an der niederländischen Stiftung SURF. In diesem Video wird ab Minute 9:20 über den Einsatz von KI bei der Metadatengenerierung gesprochen. Das Ziel des

Projektes ist es die Ergebnisse von Suchanfragen durch die Analyse des Contents zu verbessern. Dabei soll die Qualität der Metadaten durch semi-automatisches Hinzufügen verbessert werden.

- **Einschätzung:** Es konnte keine dedizierte Projektwebseite gefunden werden, auf eine Email-Anfrage wurde nicht reagiert.
- Projekt an der TU Delft (Niederlande): Über eine Anfrage bei Robert Schuwer wurde bekannt gemacht, dass es an der TU Delft ein Projekt gibt, das sich ebenfalls mit den Themen KI, Metadaten und Suche befasst. Verantwortlich ist Michiel de Jong.
  - **Einschätzung:** Eine dedizierte Projektwebseite existiert nicht. Auf eine Anfrage hin wurde mitgeteilt, dass das Projekt noch in einer sehr frühen Phase sei und noch keine Ergebnisse vorlägen.

Zusammenfassend lässt sich feststellen, dass es sehr vereinzelt Projekte gibt, die sich dem Thema KI und Metadaten im Bereich OER widmen, jedoch sind diese Themenbereiche national höchstens in einem Arbeitspaket eines Teilprojektes angesiedelt (EduArc). Auf europäischer Ebene erscheint das Projekt X5GON vielversprechend, vor allem der zu entwickelnde Service X5oerfeed. Die Anwendbarkeit und Praxistauglichkeit kann zu diesem Zeitpunkt noch nicht bewertet werden.

Wird der Blick über den expliziten Bereich OER erweitert, lassen sich weitere Projekte ausmachen, die sich ähnlichen Themen befassen. Das in Finnland für den Bibliotheksbereich entwickelte Tool [Annif](#) lässt sich beispielsweise für Indexierung und Klassifikation nutzen. Im Redaktionsbereich und Journalismusbereich wurde bspw. der [ARD-Media-Data-Hub](#) entwickelt, der KI-gestützte Metadatengenerierung und crossmediale Suchen anbietet. Allgemein scheint bei diesem Thema ein Blick über den Tellerrand von OER hinaus sinnvoll und nötig, um zu überprüfen, ob ähnliche Problemstellungen in anderen Fachbereichen Adaptionmöglichkeiten auf den (O)ER-Kontext bieten.

# Herausforderung 7: Verknüpfung der Metadaten mit externen Wissensquellen ausbauen

## *Wie können Linked Data Konzepte im OER-Umfeld sinnvoll genutzt werden?*

Der Begriff "Linked Open Data" (LOD) bezeichnet eine Community-Bestrebung, innerhalb derer Regeln und Best Practices definiert werden, um große Datensätze an offenen Metadaten maschinenlesbar zu veröffentlichen und mittels semantischer Techniken miteinander in Beziehung zu setzen.

Tim Berners-Lee definierte dazu folgende vier Regeln:

1. Benutze Uniform Resource Identifiers (URIs) als Namen für Dinge.
2. Benutze HTTP URIs, so dass diese Namen nachgeschlagen werden können.
3. Wenn jemand eine URI nachschlägt, liefere nützliche Informationen und nutze dabei die Standards RDF\* und SPARQL.
4. Füge Links zu anderen URIs bei, so dass der User weitere Dinge entdecken kann.

Diese ständig wachsende Wissensbasis wird als Linked Open Data Cloud (LOD Cloud) bezeichnet und beinhaltet aktuell über 28 Milliarden RDF-Triple aus ca. 650.000 Datenquellen.

Im OER Umfeld existieren bereits interessante LOD-Datenquellen, die mit OER Materialien verknüpft werden könnten, z.B. :

- Wikidata ist eine umfassende frei bearbeitbare Wissensdatenbank, die von der Wikimedia Foundation u.a. zur Unterstützung der verschiedenen sprachigen Wikipedia-Instanzen mit einheitlichen strukturierten Daten betrieben wird. Ein aktuelles Beispielprojekt, das auf Wikidata basiert, ist Science Stories (<http://www.sciencestories.io/>), das eindrucksvoll die Möglichkeiten zeigt.
- Der LOD-Dienst des hbz stellt die Gemeinsame Normdatei (GND) mit über 15 Millionen LOD-Datensätzen zu Personen, Körperschaften, Kongresse, Geografika, Sachschlagwörter und Werktitel zur Verfügung.
- Die Kompetenzrahmen im gesamten Bundesgebiet könnten in maschinenlesbarer Form erstellt (siehe Herausforderung 1) und als Linked Open Data veröffentlicht werden.

Um dieses Wissen der einzelnen Datenquellen zusammenzufassen, ist es zwingend erforderlich, entsprechende Klassen von Ressourcen und Properties, sowie deren Zusammenhänge zu definieren, auf deren Basis dann die einzelnen Datenquellen ihre strukturierten Daten herausgeben können.

Die Initiative [schema.org](https://schema.org) stellt so eine Ontologie für die Strukturierung von Daten auf Webseiten zur Verfügung. Im Bildungsbereich wurde der [LRMI-Standard](#) von der [Dublin Core Metadata Initiative](#) herausgegeben, um auf Basis von schema.org Bildungsressourcen beschreiben zu können.

Durch die Verknüpfung der OER Materialien mit der LOD Cloud wäre es zum Beispiel möglich, bei der Suche nach OER Materialien über eine Stadt anbieten zu können, den Suchkontext auf die entsprechende Region bzw. das jeweilige Bundesland auszuweiten, indem man die Verknüpfung von den OER Metadaten zu einer LOD Ressource verfolgt und iterativ in die Suche einbezieht. Man könnte des weiteren auch Materialien von berühmten Personen/Gebäuden aus der Region finden und so weiter und so fort.

Wenn man die Metadaten der OER Materialien als Linked Open Data veröffentlicht, dann könnte das veröffentlichte Materialien auch wiederum Teil der LOD Cloud werden. Somit wäre es dann zum Beispiel möglich, OER Materialien zu finden, die in ihren Metadaten nicht direkt den Bezug zu einer LOD Ressource aufweisen, aber Bestandteile mit dem Bezug dazu besitzen.

Entweder werden die generierten LODs von der OER Datenquelle direkt in die Website für das jeweilige OER Material eingebettet oder explizit mittels Content Negotiation zur Verfügung gestellt. Bei Einsatz von PID-Systemen könnte dies von den eingesetzten PID-Services (siehe [Herausforderung 2](#)) übernommen werden, da sie auch die Metadaten des veröffentlichten OER Materials vorhalten und diese als LOD bei der PID-Auflösung mittels Content Negotiation direkt ausgeben könnten. Eine Weiterleitung auf die jeweilige OER Datenquelle würde dann wegfallen.

# Herausforderung 8: Transparenz der OER-Produktion verbessern

## *Wie können Metadaten über die Herkunft und die Zusammensetzung von OER Materialien generiert werden?*

Ein wesentlicher Vorteil von OER ist, dass sie – aufgrund ihrer offenen Lizenzierung – wiederverwendet, angepasst, vermischt und wiederveröffentlicht werden können. Ziel ist der Aufbau einer Kultur, in dem kollaborativ an der stetigen, schrittweisen Verbesserung von Materialien gearbeitet wird.

Dieser OER-Produktionsprozess sollte transparent gestaltet und über die Metadaten jederzeit die Herkunft zurückverfolgt werden können (siehe [Herausforderung 2](#)). Bei einem OER-Remix sollten zusätzlich die Referenzen auf das wiederverwendete Material mit angegeben werden (siehe [Herausforderung 3](#)). Die manuelle Erfassung dieser **Provenance-Metadaten** ist sehr aufwendig. Man kann aber versuchen, sie im Rahmen des Produktionsprozesses automatisiert zu generieren.

**Klassische Autorensysteme** (Word etc.) erzeugen meist zur Laufzeit das endgültige Artefakt, um eine Echtdarstellung (WYSIWYG) anbieten zu können. Abhängig von dem gewählten Ausgabeformat können darin auch Metadaten enthalten sein. Die Systeme stellen meist intern eine Bearbeitungshistorie zur Verfügung, so dass entsprechende Zeitpunkte für die Erstellung / Bearbeitung in den Metadaten hinterlegt werden können. Erfolgt zuvor eine Authentifizierung, dann könnte auch personenbezogene Informationen aus dem Benutzerprofil hinzugefügt werden. Bei einem OER-Remix wird das wiederverwendete Material importiert, der Link auf die Ressource muss aber zusätzlich in die Metadaten mit einfließen.

Ebenfalls existieren **web-basierte Autorensysteme** (moodle etc.), bei denen aktuell Inhalt und Darstellung der Materialien getrennt voneinander definiert wird, um jederzeit eine Änderung des Layouts unabhängig vom Inhalt vornehmen zu können. Durch Einsatz web-basierter Editoren kann trotzdem eine WYSIWYG-Darstellung im Browser angeboten werden, im Hintergrund wird der Inhalt aber getrennt vom Layout mit Hilfe einer Auszeichnungssprache (z. B. XML) abgelegt. Für die spätere Darstellung wird der Inhalt zur Laufzeit mit Hilfe einer Layoutvorlage in das gewünschte Ausgabeformat (z.B. HTML) transformiert. Auch bei diesem Ansatz müssen die Provenance-Metadaten von den Autorensystemen zur Verfügung gestellt werden, weil diese in der Regel kein darunterliegendes Versionierungssystem nutzen. Nur beim OER-Remix genügt hier die Angabe eines stabilen Links auf die Web-Ressource, die Auflösung und Attributierung wird automatisch bei der

Transformation durchgeführt. Es existieren standardisierte Im-/Export-Formate für web-basierte Bildungsmaterialien, die auch einen Metadaten-Austausch vorsehen.<sup>9</sup>

Sehr prominent sind momentan auch **text-basierte Autorensysteme**, bei denen der Inhalt zuerst textuell mit Hilfe einer Auszeichnungssprache (z. B. LaTeX, Markdown) erfasst und nachgelagert das endgültige Material (inkl. Metadaten) generiert wird.

Bei diesem Ansatz ergeben sich viele Parallelen zu Kultur und Prozessen in der Open Source Community. Das zentrale Werkzeug bei der Verwaltung von Artefakten im Rahmen des Entwicklungsprozesses ist die Versionsverwaltung. Git ermöglicht die Versionsverwaltung für textbasierte Ressourcen durch an verteilten Orten kollaborierender Personen. Auf Git aufsetzend ist ein ganzes Ökosystem weiterer Software entstanden, z. B. zur Erleichterung und Strukturierung gemeinsamer Arbeitsprozesse. Folgende Provenance-Metadaten können mit Hilfe eines **Git-basierten Workflows** automatisch zur Verfügung gestellt werden:

- Fragen der **Attribution** (Wer hat was zu einer Ressource beigetragen?) werden bis hin zur Ebene einzelner Zeichen festgehalten, wobei dies auf die Angabe von Name und E-Mail beschränkt ist. Diese Zuweisung kann mittels digitaler GPG-Zertifikate zugesichert werden<sup>10</sup>, die in einem dezentralen Netzwerk von Vertrauensstellungen (“Web of Trust”) verwendet werden können.
- **Versionen** und deren Verknüpfungen untereinander werden erfasst: Wann wurde von wem eine Version einer Ressource erstellt und auf Basis welcher Vorgängerversion.

An der Hamburg Open Online University wird seit Jahren dieser Git-basierte Ansatz verfolgt (siehe Blog) und auch die TIB (siehe Workshop) wirbt für diesen Ansatz im OER-Produktionsprozess.

Selbstverständlich bringt dieser Git-basierte Ansatz auch einige neue Herausforderungen mit sich:

- So ist die Arbeit mit Textdateien vielen Lehrenden nicht vertraut, auch wenn z.B. Markdown eine niedrige Einstiegshürde bietet.
- Der Generierungsprozess für die endgültigen Artefakte benötigt Ressourcen und Zeit, eine Echtdarstellung (WYSIWYG) kann dadurch nicht angeboten

---

<sup>9</sup> <http://www.imsglobal.org/activity/common-cartridge>

<sup>10</sup> <https://git-scm.com/book/en/v2/Git-Tools-Signing-Your-Work>

## Aktuelle Herausforderungen im Bereich OER-Metadaten

Link zum Dokument: <https://kurzelinks.de/oer-metadaten-herausforderungen>

werden.

- Zudem sind Werkzeuge wie Git für Menschen, die nicht bereits in der Softwareentwicklung damit arbeiten, zu komplex und nicht leicht zu erlernen.
- Um zugesicherte Provenance-Metadaten zu erhalten, müssen für jede:n Akteur:in entsprechende digitale GPG-Zertifikate erstellt & verteilt und ggf. entsprechende Vertrauensstellungen aufgebaut werden.

Mögliche Lösungsansätze wären, leicht bedienbare browserbasierte Autorensysteme für Markdown/HTML zu entwickeln und die Komplexität von Git hinter einer leicht zugänglichen Oberfläche zu verbergen.