



# Ausschreibung Data Scientist\*in/Data Engineer\*in 2023-2024

Vergabe es2022jo4a

ITs.JOINTLY Projekt



# Inhalt

<b>Einleitung</b>	<b>3</b>
<b>Lose</b>	<b>4</b>
<b>Los 1: Generierung von rechtlichen und qualitativen Metadaten für Bildungsinhalte</b>	<b>4</b>
<b>Los 2: Qualitätsbewertung von Metadaten mit integriertem Zeitpunktvergleich</b>	<b>4</b>
<b>Los 3: Metadatenstrukturen zur Erfassung pädagogischer Szenarien und der automatisierten Verknüpfung mit Lernmaterialien</b>	<b>4</b>

## Einleitung

In den vergangenen Jahren hat WirLernenOnline als Corona-Nothilfeprojekt mehr als 200.000 Lerninhalte erschlossen. Ad hoc wurde eine leistungsfähige Infrastruktur für die Erschließung, Verwaltung, Kuratierung und Bereitstellung von Bildungsmaterialien aus Open Source Bausteinen geschaffen. Darüber hinaus verschlagworteten ca. 30 Redaktionen 30.000 Inhalte und stellten ca. 3.000 qualitätsgesicherte Lehrplansammlungen zusammen.

Dies ermöglicht es, für kommende Projektphasen maschinelle Verfahren für Erschließung und Kuratierung zu maximieren und manuelle Arbeiten zu minimieren. Künftige Redaktionen werden zunehmend maschinelle Verfahren überwachen und steuern, methodische Ansätze optimieren und Transferleistungen erbringen.

Die Weiterentwicklung maschineller Verfahren und methodischer Ansätze ist zeitnah erforderlich, da in einem Bildungsraum zusammengeschlossener Plattformen große Inholdemengen zusammenkommen. Diese haben, soweit überhaupt ausreichend digitalisiert, sehr heterogene Daten- und Prozesslogiken.

Im Rahmen dieses Projektes sollen die entstandenen Konzepte mittels testbarer PoCs validiert und umgesetzt werden.

Für den Bereich Datenanalysen, Entwicklung intelligenter Algorithmen und Informationsvernetzung braucht das Projektteam Unterstützung. Benötigte Leistungen werden in den Losen dargestellt, in die diese Ausschreibung aufgeteilt ist.

## Lose

Diese Ausschreibung ist in verschiedene Lose unterteilt, um die Angebotsabgabe und Auswahl der Dienstleister zu erleichtern. Beschreibungen, Leistungsverzeichnisse, Termine etc. befinden sich in den jeweiligen Losen. Die Lose sollen unabhängig voneinander bearbeitet werden.

### Los 1: Generierung von rechtlichen und qualitativen Metadaten für Bildungsinhalte

Im Rahmen dieses Loses soll ein Konzept für die maschinelle Prüfung von textlichen, visuellen und audiovisuellen Bildungsinhalten auf rechtliche und qualitative Aspekte entstehen, das nach positiver Machbarkeitsentscheidung bis Ende 2024 umgesetzt und validiert werden soll.

### Los 2: Qualitätsbewertung von Metadaten mit integriertem Zeitpunktvergleich

Im Rahmen dieses Loses sollen bestehende Konzepte zur qualitativen Beurteilung von Metadaten (z.B. Beschreibungstexten) auf ihre Vollständigkeit und Umsetzbarkeit hin untersucht und ggf. dahin erweitert werden, mit dem Ziel, einen Proof of Concept für eine automatisierte Bewertung zu erstellen. Nach positiver Machbarkeitsentscheidung soll es in der zweiten Phase bis Ende 2024 umgesetzt und validiert werden.

### Los 3: Metadatenstrukturen zur Erfassung pädagogischer Szenarien und der automatisierten Verknüpfung mit Lernmaterialien

Im Rahmen dieses Loses soll ein Konzept für die Entwicklung eines skalierbaren, flexiblen und nachhaltigen Metadatenmodells pädagogischer Szenarien entstehen sowie die automatische Verknüpfung der Szenarien mit Lerninhalten konzipiert und erprobt werden. Es ist beabsichtigt, ausführbare Proof of Concepts (PoCs) zu entwickeln, um die grundsätzliche Machbarkeit zu demonstrieren und die erforderlichen Anpassungen unserer Datenprofile für Lerninhalte zu entwerfen und zu testen.