

OER als Schlüssel zur Data Literacy

Lessons Learned aus dem E-Learning-Projekt
„MaDaLi²“ der Universität Mannheim



MADALI²
@UniMannheim



Agenda

1. Projektvorstellung MaDaLi²
2. Entwicklung des E-Learning-Kurses
3. OER-Recherche
4. Prototypisierung
5. Beta-Version
6. Fazit

Projektvorstellung MaDaLi²

Projektvorstellung MaDaLi²



- **Universitätsweites** Kooperationsprojekt
 - Hochschuldidaktik (ZLL)
 - Universitätsbibliothek (Forschungsdatenzentrum)
 - Lehrstuhl für BWL | Quantitatives Marketing
 - Lehrstuhl für Medien- und Kommunikationswissenschaft
 - Lehrstuhl für Psycholinguistik
- **Projektziel:** Data-Literacy-Angebote für alle Studierenden der Universität
- **Projektlaufzeit:** 2024–2025



<https://www.uni-mannheim.de/universitaet/lehre/lehrprojekte/data-literacy/>

Projektvorstellung MaDaLi²



- **Projektansatz:**
 - Fachübergreifender Grundlagenkurs als **E-Learning-Angebot**
 - Extracurricularer Kurs mit Möglichkeit zur **Integration in Lehrveranstaltungen**
 - Fokus auf bereits existierenden **Open Educational Resources (OER)**
 - Eventuelle Lücken durch Eigenproduktion schließen
- **Limitierungen:** Ressourcen (Zeit, Personal)



<https://www.uni-mannheim.de/universitaet/lehre/lehrprojekte/data-literacy/>

Projektvorstellung MaDaLi²



Weshalb Fokus auf OER?

- **Vorrecherche:** hochwertige OER vorhanden
- kaum zeitliche Ressourcen für **umfangreiche Produktion**
- **Open Science** und **FDM** ernst nehmen

Herausforderung

- Steht **Heterogenität** der OER einem kohärenten E-Learning-Angebot entgegen?



<https://www.uni-mannheim.de/universitaet/lehre/lehrprojekte/data-literacy/>

Projektvor

Wir nehmen das O in OER ernst



Weshalb Fok

- Vorrecher
- kaum zeitl
- Produktio
- Open Scie

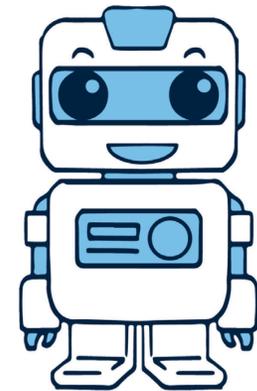
Herausforde

- Steht Hete
- E-Learning



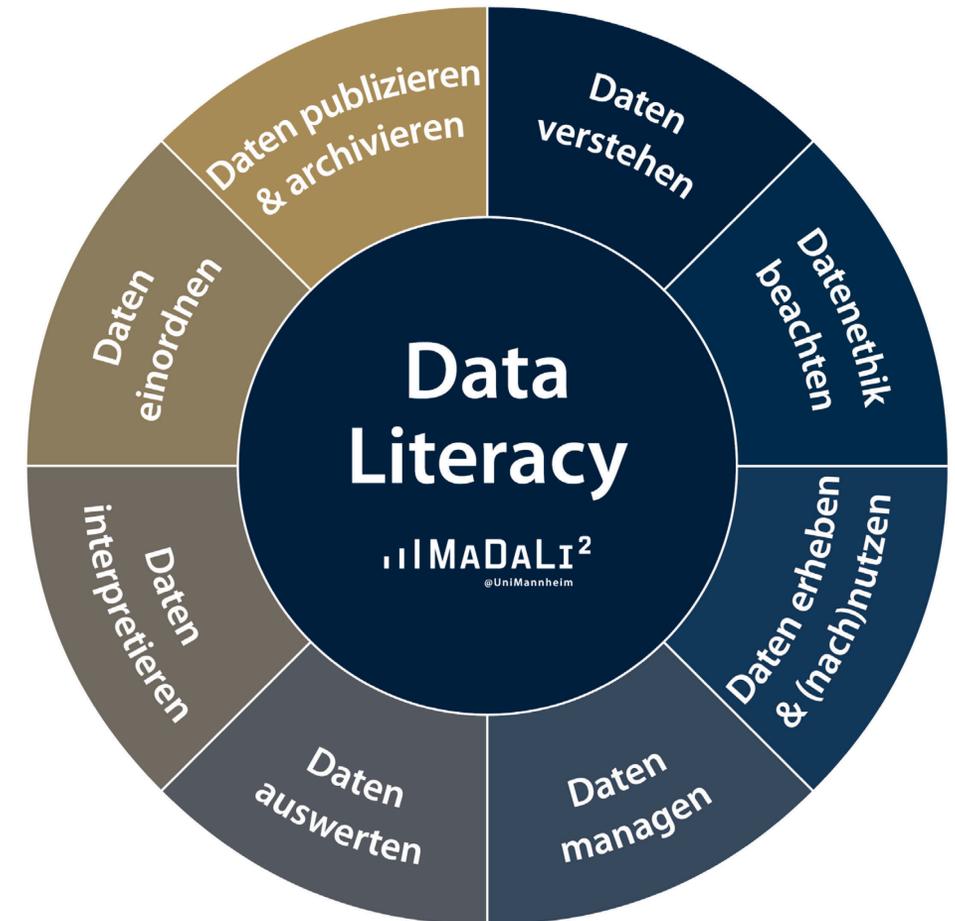
[w.uni-
de/universitaet/lehre/lehrproj
iteracy/](https://www.uni-mannheim.de/universitaet/lehre/lehrproj-iteracy/)

Entwicklung des E-Learning-Kurses



Entwicklung des E-Learning-Kurses

- **Konzeptuelle Grundlage:**
 - Data-Literacy-Charta des Stifterverbands für die Deutsche Wissenschaft e.V.
 - Data-Literacy-Frameworks (TH Köln)
 - Future Skills: Ansätze zur Vermittlung von Data Literacy in der Hochschulbildung (Hochschulforum Digitalisierung)
- **8 inhaltliche Module** zur Vermittlung von Datenkompetenzen
- **Arbeitsaufwand** für Studierende: ca. 30 h



Entwicklung des E-Learning-Kurses

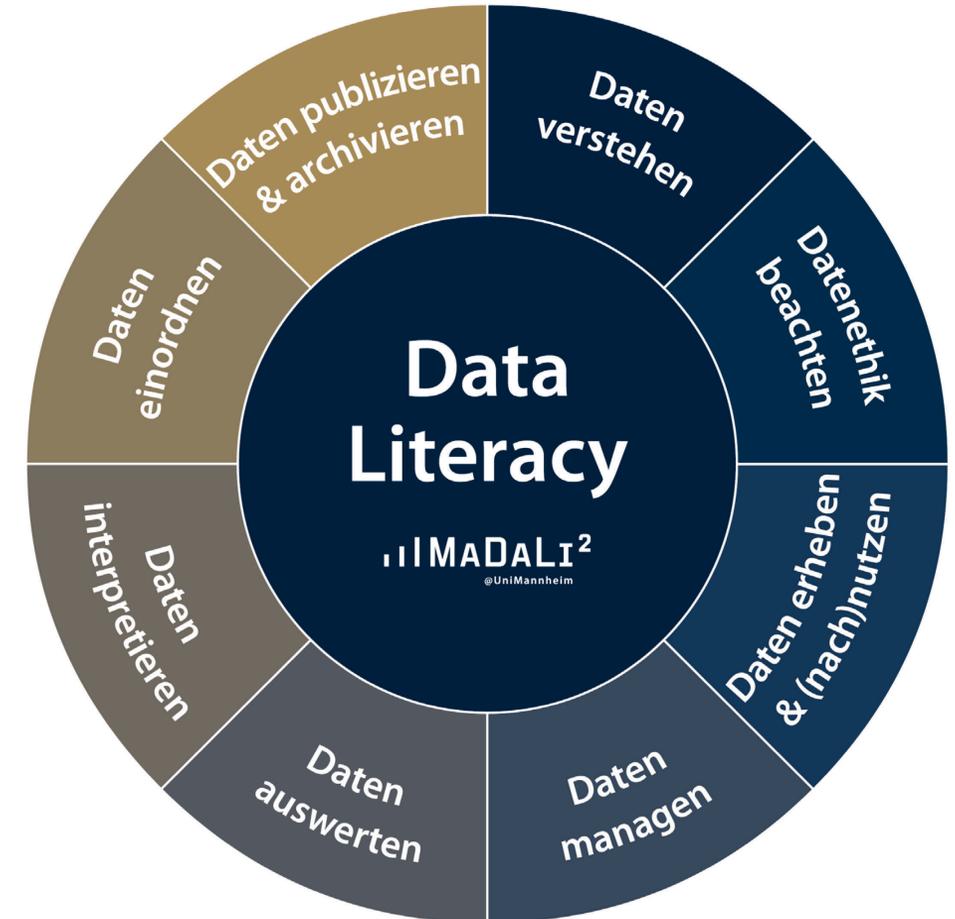
- **Lernzielentwicklung** je Modul auf Basis von Data-Literacy-Frameworks

Modul 1: Datenkultur etablieren	Lernziele
	Die Studierenden können definieren, was Data Literacy ist und sie können Bereiche, die unter Data Literacy fallen, benennen.
	Die Studierenden können erläutern, was Daten sind und Beispiele nennen.
	Die Studierenden können benennen, wo ihnen im Alltag Daten begegnen.
	Die Studierenden können Tabellen und Visualisierungen aus dem Alltag korrekt lesen.
	Die Studierenden können grob den Unterschied zwischen einer qualitativen und einer quantitativen Untersuchung erläutern.
	Die Studierenden können die verschiedenen Data-Literacy-Kompetenzen nennen und den (Forschungs-) Datenlebenszyklus beschreiben.

Auszug aus dem Curriculum des E-Learning-Kurses

Entwicklung des E-Learning-Kurses

- Module **individuell** zu absolvieren
- Module in Lehre **integrierbar**
- Bereitstellung des Kurses über die universitäre **E-Learning-Plattform ILIAS**
 - freie Open-Source-Lernplattform
 - Prüfungsmöglichkeiten



Entwicklung des E-Learning-Kurses

- H5P („HTML5 Package“) ist ein **freies** (MIT Lizenz), **quelloffenes** Framework zur Erstellung interaktiver Webinhalte
- H5P „**Interactive Book**“ als Gestaltungselement:
 - Gliederung der Modulinhalte nach **Kapiteln**
 - Große Anzahl frei wählbarer **Inhaltstypen**
 - Einfache Einbindung **multimedialer** Elemente: Bilder, Videos, Präsentationen etc.
 - Vorgefertigte Sets für **Lernaufgaben**: Lückentexte, Single & Multiple Choice Fragen, Memory Game etc.



Bild: <https://h5p.org/>



Mark the Words
Memory Game
Multiple Choice
Question Set
Single Choice Set
Summary
Timeline
True/False Question

OER-Recherche

OER-Recherche: Plattformen

- **Umsetzungsphase:** Recherche nach qualitativ hochwertigen OER:
 - **ZOERR** → <https://www.oerbw.de>: Zentrales OER-Repository der Hochschulen in Baden-Württemberg
 - **ORCA** → <https://www.orca.nrw>: Open Resources Campus NRW
 - **OERSI** → <https://oersi.org/resources>: Open Educational Resources Search Index
 - **Zenodo** → <https://zenodo.org>: Fachübergreifendes Open-Access-Repository
 - **TIB AV-Portal** → <https://av.tib.eu>: offene, nicht-kommerzielle Plattform für wissenschaftliche Videos
- **Sprache** der OER: Deutsch

OER-Recherche: Allgemeines

- **OER zu Grundlagen** im Fokus der Recherche
- Recherche erfolgte auf Basis einer **Schlagwortsuche** auf den identifizierten Plattformen
- **Formate:** Videos, Grafiken, Foliensätze
- **Lizenzen:**
 - Public Domain
 - CC BY
 - CC BY-SA

OER-Recherche: Videoformate

- Auf OER-Plattformen große Anzahl von **Vorlesungsmitschnitten**, die sich durch Länge und Sprechmodus nur bedingt für Selbstlernkurs eignen
- Bevorzugte **Videolänge**: 2 bis 10 Minuten
- **Ergebnis**:
 - Viele thematische **Einzelvideos**
 - Aber auch ganze **Vidoreihen** unter **freien Lizenzen** veröffentlicht



TIB AV-PORTAL

Was ist Data Literacy?

Datenkompetenz



© 26 Zitieren Teilen

Persike, Malte: Was ist Data Literacy?. data.RWTH, Rheinisch-Westfälische Technische Hochschule (RWTH) Aachen et al., 2021. <https://doi.org/10.5446/63659>

OER-Recherche: Lessons Learned

- Große **Anzahl** verfügbarer OER
- Aber auch große **Heterogenität** der Ressourcen
 - teils Suche nach der „Nadel im Heuhaufen“
 - Fokus der jeweiligen OER **sehr unterschiedlich**: auf eine bestimmte Lehrveranstaltung bezogen oder sehr allgemein
 - **uneinheitliche Verschlagwortung** der Ressourcen → Suche teils problematisch
- Häufig **überschneiden** sich Angebote auf verschiedenen Plattformen
- Lizenzen meist offen gewählt (CC BY, CC BY-SA)

OER-Recherche: Lessons Learned

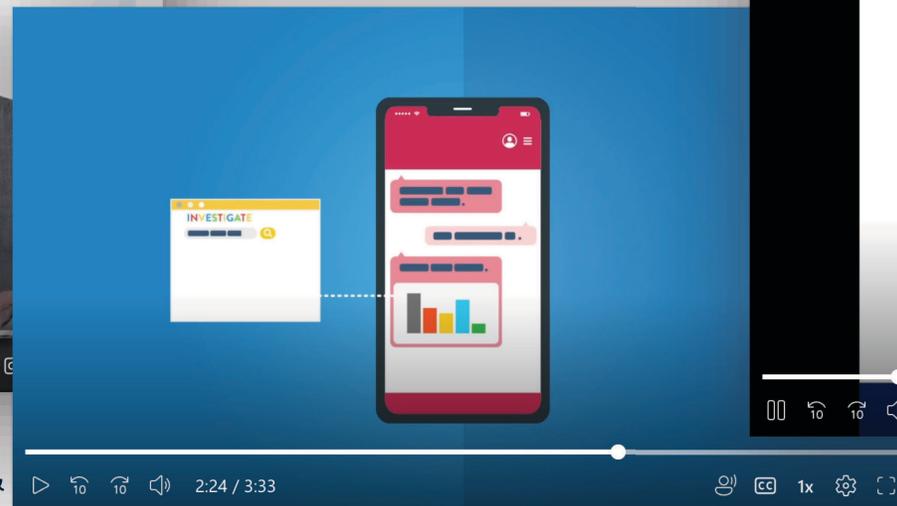
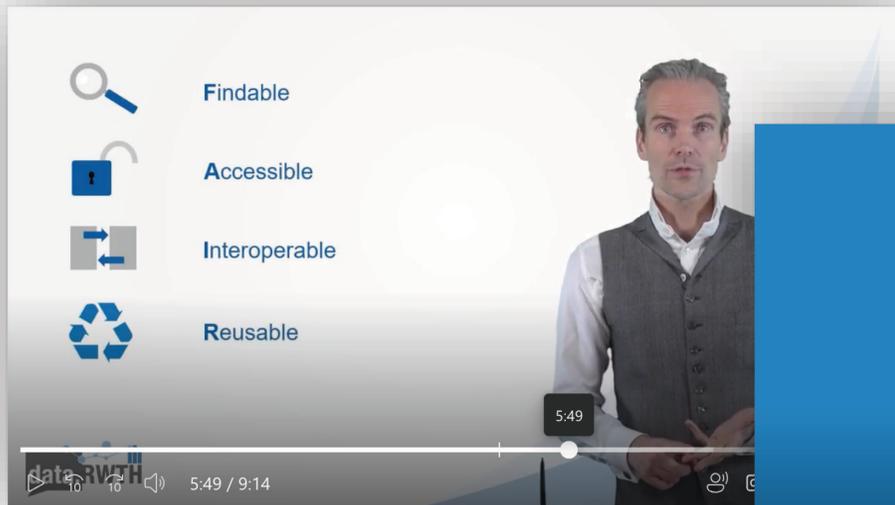
- Identifikation **inhaltlicher Lücken**: Kaum OER für geisteswissenschaftliche Disziplinen mit Fokus auf Datenkompetenz
- Neben OER-Plattformen wird auch **YouTube** für die Publikation von Materialien verwendet:
 - Trotz Vergabe von Public Domain und CC-Lizenzen erschwert YouTube die Nachnutzung, da **Videodownloads nur mit YouTube Premium** möglich sind
 - <https://support.google.com/youtube/answer/2797468>



Prototypisierung

Prototypisierung: Herausforderungen

- Recherchierte OER sind **optisch** und **strukturell heterogen**:
 - Folien-Sets mit Voice-Over, Talking-Head-Videos, Animationen, etc.

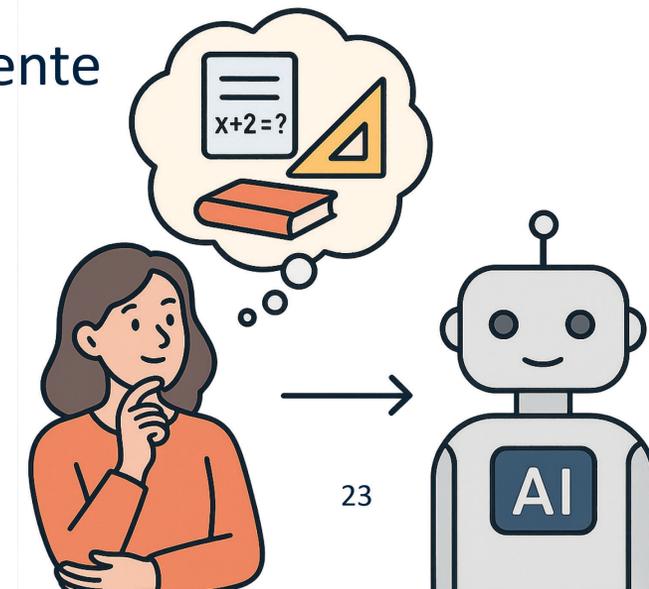


Prototypisierung: Herausforderungen

- Heterogene OER müssen mit **Kontextinformationen** (z. B. institutionelle Zugehörigkeit) gerahmt werden
- Einzelvideos aus Videoreihen zu **extrahieren** und in neuen Zusammenhang zu stellen ist **aufwändig**
- **Bearbeitung** von Videos teils notwendig (sofern es Lizenz erlaubt) – entsprechendes Know-How (Videoschnitt) erforderlich
- **Quellen-** und **Lizenzangaben** müssen detailliert dokumentiert werden!

Prototypisierung: Einbindung der OER

- **Verbindungselemente** in H5P „Interactive Books“ schaffen inhaltliche Rahmung:
 - Einheitlicher Modulaufbau: Lernziele zum Einstieg, Quellen zum Abschluss
 - Einführende Text für Videos
- Erstellung von **Transkripten** und **Untertiteln**
- **Aufgabengestaltung** auf Basis der „Interactive-Book“-Elemente
 - GenAI kann hier produktiv unterstützen!
- Produktion **neuer Inhalte** bei inhaltlichen Lücken
 - vor allem Grafiken, Texte und kurze Präsentationen



Prototypisierung: Evaluation

- Evaluation des Prototyps durch Befragung wissenschaftlicher Hilfskräfte mittels **qualitativer Leitfadeninterviews**
- **Evaluationsergebnis:**
 - Heterogene OER wurden nicht als hinderlich oder störend empfunden
 - Kritik an Aufgaben (besonders Lückentexte)
 - Modulprototyp wurde als einheitlich und stimmig wahrgenommen
- Auf Basis der Evaluation: **Anpassung der Lernaufgaben**

Beta-Version

Beta-Version

- Launch der Beta-Version des E-Learning-Kurses im **Februar 2025**
- **Großes Interesse** seitens der Studierenden (und Lehrenden)
 - Stand Mai 2025: ca. 250 Personen eingeschrieben
- Integration einzelner Module in **Lehrveranstaltungen** (Blended Learning)
 - Anglistische Linguistik
 - Medien- und Kommunikationswissenschaften
- **Evaluation** der Beta-Version durch kurze Online-Surveys



Modul 1: Datenkultur eta...

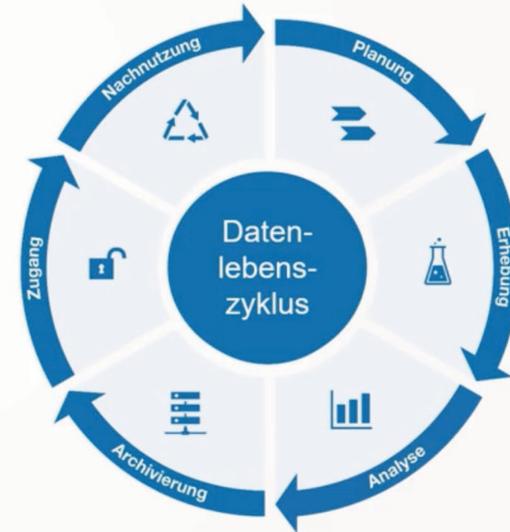
- ▶ Einstieg ○
- ▶ Was ist Data Literacy? ○
- ▶ Was sind Daten? ○
- ▶ Daten im Alltag ○
- ▼ Der Datenlebenszyk... ○
 - Video: Der Datenlebenszykl...
 - Lernquiz
 - Datenlebenszyklus (Fragen...
 - FAIR Data (Lückentext)
- ▶ Forschungsdaten - ... ○
- ▶ ...tion ○
- ▶ ...lus ○

Video: Der Datenlebenszyklus und FAIR Data



Der Datenlebenszyklus ist ein Modell, das die verschiedenen Phasen der Planung, Erhebung, Analyse, Archivierung, Zugänglichkeit und Nachnutzung, die Daten durchlaufen können, beschreibt.

(vgl. RWTH 2021)



In diesem Video erläutert Malte Persike von der RWTH Aachen den **Datenlebenszyklus** und die sogenannten **FAIR-Prinzipien**. Sie lernen die Phasen des Datenlebenszyklus kennen und erfahren, wieso es sinnvoll ist, **Daten auch längerfristig nutzbar** zu machen.

> **Videotranskript zum Nachlesen**



Modul 6: Daten interpreti...

▶ Einstieg ●

▼ Verzerrungen in den... ○

Überblick

Video: Verzerrungen in For...

Video: Verzerrung der Stich...

Exkurs: Video - Machine Le...

Video: Publication Bias

Video: Questionable Resea...

Lernquiz

○ Verzerrungen in den Daten ...

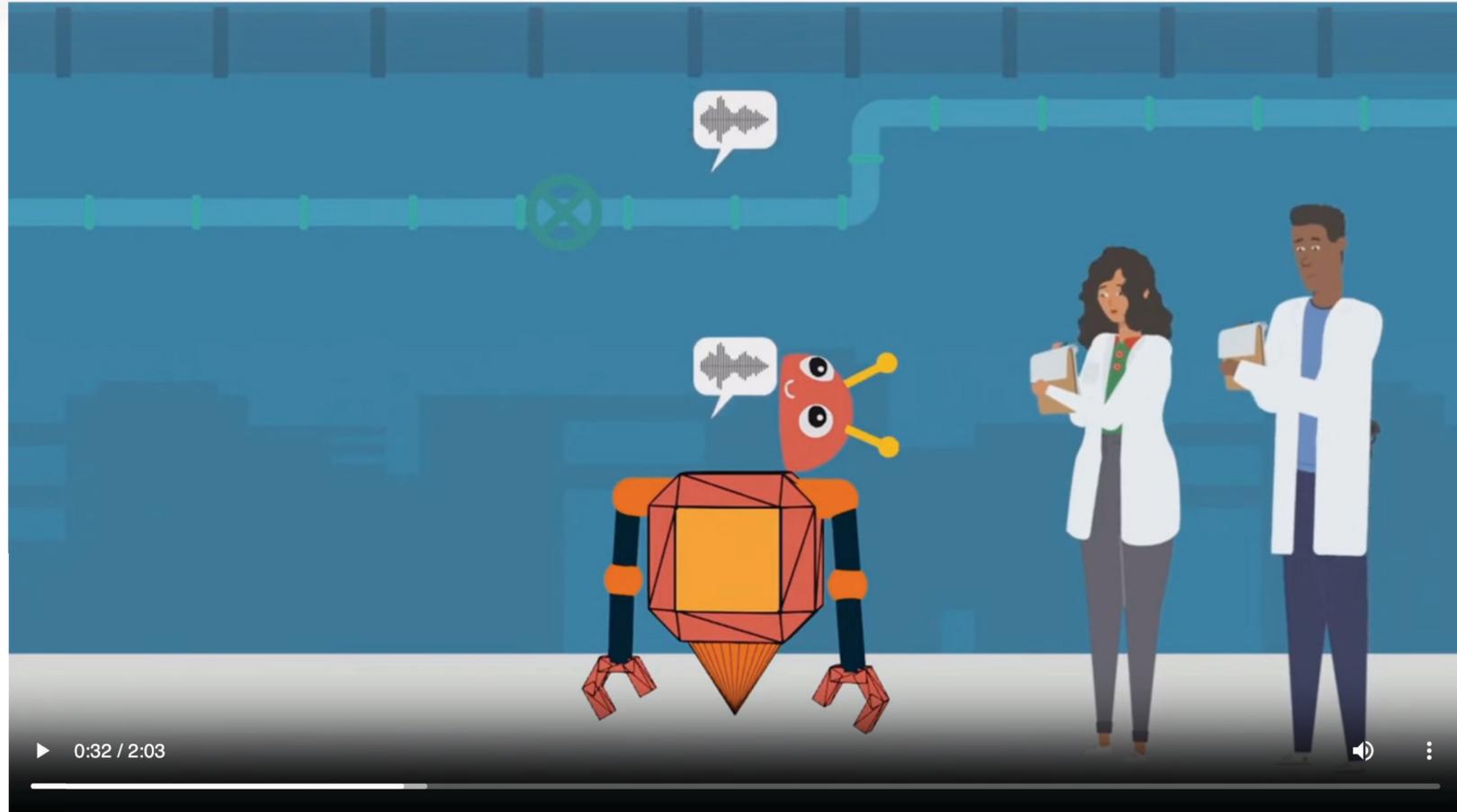
▶ Verzerrungen bei de... ○

▶ Fallstricke bei der I... ○

▶ Anwendungsbeispi... ○

on ○

Beta-
Version



▶ 0:32 / 2:03

▼ Videotranskript zum Nachlesen

Für maschinelles Lernen benötigt der Computer Trainingsdaten. Aus diesen Daten lernt der Computer. Es ist sehr wichtig, dass der Computer mit vielen und hochwertigen Daten lernt. Doch woher stammen diese Trainingsdaten?

Mit welchen Daten ein Computer lernt, entscheiden wir Menschen. Zum Beispiel wählen Linguistinnen und Linguisten Sprachaufnahmen aus und stellen diese dem Computer bereit. Daraus kann der Computer lernen, Sprachbefehle zu verstehen. Die Qualität der Trainingsdaten ist sehr wichtig. Lernt der Computer beispielsweise nur mit Sprachaufnahmen von Frauen, sind die Trainingsdaten unausgewogen. Womöglich kann der Computer dann Männer nicht verstehen. Durch unausgewogene Trainingsdaten entsteht eine Verzerrung, auch Bias genannt. Einige Szenarien

Modul 3: Daten erheben ...

- ▶ Einstieg ●
- ▶ Warum empirisch f... ○
- ▶ Quantitative und q... ○
- ▶ Quantitative und q... ○
- ▶ Methoden der Date... ○
- ▶ Forschungsdaten (... ●
- ▶ Forschungsdaten r... ○
- ▼ Forschungsdaten r... ○

Video: Wie finde ich ein p...

Aufgabe: Recherchieren ...

re3data.org in der Recherche (Si...

Forschungsdaten r...
g...

Beta-
Version

 Zusammenfassung und
Einsenden

Um die Inhalte des letzten Kapitels zu vertiefen, führt Sie dieses kurze Video des Forschungsdatenmanagements Bayern in die Recherche geeigneter Repositorien mit [re3data.org](https://www.re3data.org) ein.

> **Videotranskript zum Nachlesen**



Aufgabe: Recherchieren mit re3data.org

Die folgende Aufgabe dient der praktischen Vertiefung des erworbenen Wissens. Folgen Sie den aufgelisteten Schritten und beantworten Sie die nachfolgenden Fragen.

1. Gehen Sie auf **re3data.org**: <https://www.re3data.org/>.
2. Grenzen Sie die Suche ein, indem Sie in der Kategorie „**Subjects**“ nach „**Developmental and Educational Psychology**“ filtern.
3. Grenzen Sie die Suche ein weiteres Mal ein, indem Sie im Bereich „**Countries**“ nach „**Germany**“ filtern.

Welche Repositorien werden Ihnen angezeigt? 1/4

Databrary und Psi Open Data

NEMAR und Edmond

Psychological Data Service (PDS) und DataCamp



Modul 8: Daten publiziere...

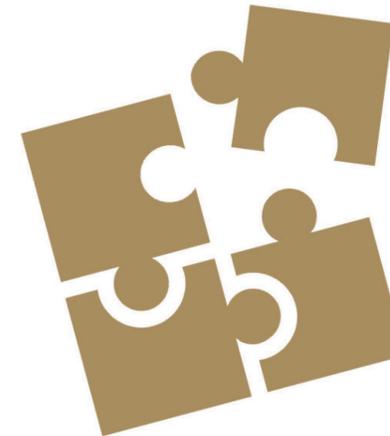
- ▶ Einstieg ●
- ▶ Die FAIR-Prinzipien ○
- ▶ Open Knowledge ○
- ▶ Open Access ○
- ▶ Open Science ○
- ▼ Exkurs: Digitale Lan... ○
 - Digitale Langzeitarchivierun...
 - Video: Die Langzeitarchivie...
 - LZA: Gründe und Strategien
- ▶ NFDI – Nationale Fo... ○
- E... on ○

> Video-Transkript zum Nachlesen

LZA: Gründe und Strategien

4 Gründe für die digitale Langzeitarchivierung

- Forschungsdaten (auch Software etc.) sollen **langfristig verfügbar** sein
- **Beschädigte** oder **veraltete Datenträger** (Festplatten, DVDs, USB-Sticks etc.) führen zu irreversiblen Datenverlust
- **Veraltete Dateiformate** und **Software** lassen keine einwandfreie Darstellung und Nachnutzung von Daten und Software zu
- Daten und Software sollen über **technologischen** und **organisatorischen Wandel** hinaus erhalten bleiben



Beta-Version: Nächste Schritte



- **Übersetzung** der Beta-Version ins **Englische**:
 - Übersetzung aller Texte, Videos, Foliensätze und Aufgaben mit Übersetzungsstelle der Universität und KI-Tools (z. B. DeepL Pro)
 - Klärung von Lizenzfragen und Kontaktaufnahme mit (fast) allen Urheber*innen
 - Wo möglich: **Übersetzung** und Bearbeitung **der Quelldaten**
 - Ressourcenintensiv! → Übersetzung von Präsentationen, Austausch von Links etc.
 - Englische **Voice-Over für alle Videos** (Text-to-Speech):
 - Marktsichtung: kommerzielle Anbieter und kostenlose Open-Source-Tools
 - Entscheidung für **kokoro** (kostenfreies Open-Source-Tools):
<https://github.com/hexgrad/kokoro>

Beta-Version: Nächste Schritte

- Launch des **finalen E-Learning-Kurses** (DE & EN) im HWS 2025/26
- Planung eines **Zertifikatskurses** (vsl. Start im HWS 2025/26)
- Konzeptionierung weiterer Module → **AI Literacy**
- **Publikation** aller Module als H5P „Interactive Books“ auf Zenodo / ZOERR
 - Lizenz: „as open as possible“

Fazit

Fazit

- **Nachnutzung** und **Integration** von OER in universitären E-Learning-Angeboten ist erfolgversprechend
- Bei entsprechender **Rahmung** stellt Heterogenität der OER keinen Nachteil dar
 - Kohärenter Kursaufbau notwendig
 - Wiederkehrende Inhaltsblöcke sorgen für Übersichtlichkeit
- Je **offener** die Lizenz, desto einfacher sind OER nachnutzbar
 - *CC BY-NC-ND* schränkt ein!
- **Aufgabensets** bisher nicht als OER veröffentlicht: Bedarf adressieren!

Vielen Dank für Ihre Aufmerksamkeit!

Dr. Leonie Bröcher: leonie.broecher@uni-mannheim.de

Thomas Schmidt: thomas.schmidt@uni-mannheim.de



Literatur

- Echtenbruck, M. M., Fühles-Ubach, S., Naujoks, B., & Kaliva, E. (2025). A Data Literacy Competence Model for Higher Education and Research (No. arXiv:2504.15690). arXiv. <https://doi.org/10.48550/arXiv.2504.15690>
- Heidrich, J., Bauer, P., & Krupka, D. (2018). Future Skills: Ansätze zur Vermittlung von Data Literacy in der Hochschulbildung.
- hexgrad. (2025). kokoro [GitHub Repository]. <https://github.com/hexgrad/kokoro>
- Lang, M., Navarro Bullock, B., Butscher, R., Gadatsch, A., Haag, B., Hummel, O., Karg, S., Klingenberg, C., Schwarz, O., Weber, K., Zimmermann, R., & Carl Hanser Verlag Verlag. (2023). Datenkompetenz: Daten erfolgreich nutzen. Hanser. <https://doi.org/10.3139/9783446477438>
- Schüller, K., Koch, H. & Rampelt F. (2021). Data-Literacy-Charta. Version 1.2. Berlin: Stifterverband.