Die Vermittlung von Best Practices zur Messung von Datenqualität in den quantitativen Sozialwissenschaften

Julian Dehne [1], Jessica Daikeler [2], Henning Silber [3], Beatrice Rammstedt [4], Florian Keusch [5], Frauke Kreuter [6], Johannes Blumenberg [7], Clemens Lechner [8], Lena Römer [9], David Schoch [10], Pascal Siegers [11], Stefan Jünger [12] und Katrin Weller [13]

Abstract: Im Kompetenzzentrum Datenqualität in den Sozialwissenschaften (KODAQS) wird ein Ansatz zur Vermittlung von Best Practices zur Messung von Datenqualität in den quantitativen Sozialwissenschaften erarbeitet und umgesetzt. Durch das entwickelte Curriculum und Modellierung eines Lehr-Lernprozesses für den Kompetenzerwerb von Methoden und Wissen zur Beurteilung von Datenqualität liefert das KODAQS Projekt einen Beitrag zum Aufbau einer Fachdidaktik und fördert damit die wissenschaftliche Qualität datengetriebener Forschung. Erste Erfahrungen, Desiderata und geplante weitere Forschung werden zur Diskussion gestellt.

Keywords: Datenqualitätsmessung, Lehr-Lernkonzept, Action Research, Praxisbeitrag, Datenqualität, Sozialwissenschaften

[1] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, julian.dehne@gesis.org, https://orcid.org/0000-0001-9265-9619
[2] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, jessica.daikeler@gesis.org, https://orcid.org/0000-0002-4879-8344
[3] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, henning.silber@gesis.org, https://orcid.org/0000-0002-3568-3257
[4] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, beatrice.rammstedt@gesis.org, https://orcid.org/0000-0002-2105-7271
[5] Universität Mannheim, A5, 6 Gebäude B, 68159 Mannheim, f.keusch@uni-mannheim.de, https://orcid.org/0000-0003-1002-4092
[6] LMU München, Ludwigstr. 33 , 80539 München, frauke.kreuter@stat.uni-muenchen.de, https://orcid.org/0000-0002-7339-2645
[7] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, johannes.blumenberg@gesis.org, https://orcid.org/0000-0003-0943-2283
[8] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, clemens.lechner@gesis.org, https://orcid.org/0000-0003-3053-8701
[9] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, lena.roemer@gesis.org, https://orcid.org/0000-0002-5885-4426
[10] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, david.schoch@gesis.org, https://orcid.org/0000-0003-2952-4812
[11] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, pascal.siegers@gesis.org, https://orcid.org/0000-0001-7899-6045
[12] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, stefan.juenger@gesis.org, https://orcid.org/0000-0001-8100-7957
[13] GESIS - Leibniz-Institut für Sozialwissenschaften, B6, 4-5, 68159 Mannheim, katrin.weller@gesis.org, https://orcid.org/0000-0003-3799-1146

doi:10.18420/inf2024_193

1 Einleitung

Es gibt vier zentrale Trends, durch die messbare Datenqualität in der sozialwissenschaftlichen Forschung zu einem Hauptanliegen geworden ist: (1) Durch die Replikationskrise [Bake16] ist sowohl das Forschungsdatenmanagement [Deut00, HaSo16] als auch die Forschungsdatenqualität in den Fokus der Aufmerksamkeit gerückt. (2) Parallel dazu hat die technologische Wende zur Entwicklung der Fächer Datenwissenschaften und computerbasierte Sozialwissenschaften geführt [Br23; FKR21; KBP15]. Dadurch ist sowohl die Fülle der verfügbaren Daten als auch deren automatisierte und skalierbare Bewertung als Lerngegenstand relevanter geworden. (3) Gleichzeitig nimmt die Bereitschaft zur Teilnahme an Befragungen, einem Hauptinstrument der Sozialwissenschaften, immer weiter ab, während parallel dazu neue Datenquellen, wie digitale Verhaltensdaten, für Sozialwissenschaftler erschlossen werden [Da24; FBD23]. (4) Zuletzt hat das nationale Interesse an der Archivierung und dem Teilen von Daten für den Forschungsstandort Deutschland zugenommen, was sich beispielsweise in der groß angelegten Förderung der Nationalen Forschungsdateninfrastruktur [BM23; Lö19] niedergeschlagen hat.

Aber was bedeutet Datenqualität für moderne sozialwissenschaftliche Daten? Im Allgemeinen bezieht sich Datenqualität auf das Ausmaß, in dem ein Satz inhärenter Merkmale von Daten [IS20] die Anforderungen der beabsichtigten betrieblichen Entscheidungsfindung und anderer potenzieller Verwendungen erfüllt (Herzog et al., 2007). In den Sozialwissenschaften wird Datenqualität oft in sogenannten Datenqualitäts- oder Fehlerrahmenwerken diskutiert [ABK20; GL10; Se21]) und als ein vielschichtiges Konzept mit vielen relevanten Dimensionen betrachtet [HSW07] . Während die Berechnung von Messfehlern, Stichprobenverzerrungen und anderer Gütekriterien für quantitative Daten schon länger beforscht wird, steckt eine umfassende Fachdidaktik zu Datenqualität inbesonderer neuer Datenquellen noch in den Anfängen. Einen Startpunkt liefert die systematische Übersichtsarbeit zu Datenqualitätskonzepten von Daikeler et al. [Da24] die etablierten Konzepte und Dimensionen der Umfrageforschung als Basis für die Datenqualität neuer Datentypen nutzt. Hier trägt das vom Bundesministerium für Bildung und Forschung (BMBF) im Förderprogramm Äufbau von Datenkompetenzzentren in der Wissenschaft"geförderte Projekt „Kompetenzzentrum Datenqualität in den Sozialwissenschaften“ (KODAQS) dazu bei, die bestehende Lücke zu schließen. Hierbei ist KODAQS eines von zehn geförderten Datenkompetenzzentren und das einzige welches sich ausschließlich mit sozialwissenschaftlichen Daten beschäftigt. Die in KODAQS aufgebauten Kompetenzen zum Kompetenzaufbau sollen hierbei nach der dreijährigen Laufzeit verstetigt werden.

KODAQS ist ein Verbundprojekt von GESIS, der Universität Mannheim und der Ludwig-Maximilians-Universität München. Es verfolgt den Zweck, die einschlägige Wissenschaftsgemeinschaft der Sozialwissenschaftler und damit verbundene Disziplinen für die Qualität sozialwissenschaftlicher Daten zu sensibilisieren und sie in der Datenqualitätsberechnung und - korrektur zu unterstützen. Primäre Zielgruppe sind Graduierte und die Angeboite adressieren Forscherinnen und Forscher vor allem bei der Sekundärdatennutzung. Hierzu

ist KODAQS ein Lern- Vernetzung- und Forschungsort. Abbildung 1 zeigt die Projektstruktur. Der Lernort KODAQS umfasst die KODAQS Academy und die KODAQS Toolbox. In diesen Infrastruktuten wird das Konzept der Datenqualität vermittelt und anhand der Indikatorik vermittelt werden. In der Data Quality Academy werden Forschende in innovativen digitalen und hybriden Lernformaten Methoden zur Bestimmung der Qualität verschiedener Datentypen bedarforientiert im Hinblick auf ihre eigenen Fragestellungen erlernen und diese an praktischen Problemstellungen aus der eigenen Forschung anwenden. Diese Bedarfsorientierung wird insbesondere durch die Möglichkeit vorangetrieben sich je nach gewähltem Datenschwerpunkt - Umfragedaten, digitale Verhaltensdaten oder verlinkte Daten, den eigenen Syllabus individuell zusammenzustellen.

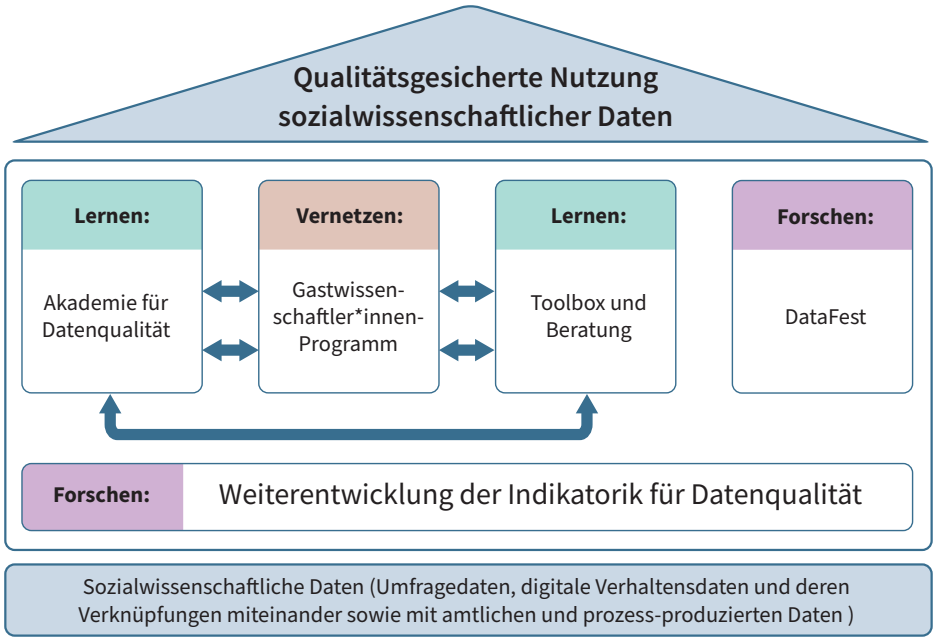

Abb. 1: Übersicht über die Projektstruktur des KODAQS Projektes

Ziel des Projekts ist es junge Forschende aber auch fortgeschrittene Wissenschaftler*innen zu befähigen, das Thema Datenqualität als Multiplikator*innen in die Hochschullehre zu tragen. Als breiteres und niederschwelligeres Angebot wird – ergänzend zur Academy – eine digitale Data Quality Toolbox, genauer eine Online Plattform mit einschlägigen Hilfestellung zur Bestimmung und Verbesserung von Daten entwickelt. die der gesamten wissenschaftlichen Community erlaubt, sich eigenständig und angeleitet Datenkompetenzen anzueignen und in transparenter und standarisierter Form die Qualität von zu untersuchenden Daten zu bestimmen und zu beurteilen. Hierfür enthält die Data Quality Toolbox nachnutzbaren Analysecode, flankiert von konkreten Anwendungsbeispielen und Beispieldatensätzen, erklärenden Videos und einem Forum, in dem sich Nutzende austauschen

und durch KODAQS-Mitarbeitende unterstützt werden können. Analysecodes, die während der Trainings der Academy auf Basis spezifischer Datensätze entwickelt werden, werden ebenfalls über die Data Quality Toolbox geteilt und zur Nachnutzung zur Verfügung gestellt. Als Hilfe zur eigenständigen Beurteilung der Qualität von Forschungsdaten werden Referenzwerte und Interpretationshilfen sowie kurze Selbsttests zur automatisierten Überprüfung des Kenntnisstands zur Verfügung gestellt. Die Toolbox wird interaktiv gestaltet, sodass Forschende direkt bei der Berechnung angeleitet werden. Schließlich wird ein individuelles, persönliches Beratungsangebot zu Datenqualität geschaffen, das Forschenden ermöglicht, zu spezifischen Fragen der Datenqualität in Bezug auf ihre eigene Forschung unterstützt zu werden. Abgerundet wird KODAQS durch einen Vernetzungsort mit einem Gastwissenschaftler*innenprogramm, das den internationalen Austausch fördert. Dazu kommen Forschungsangebote mit den sogenannten "DataFests", also Hackathons mit Fokus auf die Verbesserung der Datenqualität, sowie Grundlagenforschung zur Weiterentwicklung der Datenqualitätsindikatoren. Im Folgenden wird das Lehr-Lernkonzept von KODAQS als Lernort in die bestehende fachdidaktische und allgemeine Forschung eingeordnet. Weiterhin werden pädagogische Designentscheidung erörtert und das Curriculum der Academy vorgestellt. Abschließend werden Erfahrungsergebnisse und eine empirische Konzeptvalidierung vorgestellt.

2 Stand der Forschung

Eine Pädagogik der Vermittlung von Datenqualitätsstandards in den quantitativen Sozialwissenschaften lässt sich zwischen den Fachdidaktiken der angrenzenden Fächer Informatik, Datenwissenschaften, Statistik und Methodenlehre für die Sozialwissenschaften einordnen. Da die Datenwissenschaften bereits eine Verbindung zwischen den anderen Fachbereichen darstellen, wird auf diese zuerst eingegangen. Die sich entwickelnde Pädagogik in Data Science verläuft von der Fachentwicklung her ähnliche Bahnen wie in naheliegenden Fächern. Gemeinsam ist die Wurzel in traditionellen Fächern (Informatik basierte zum Beispiel auf der Mathematik) und die provisorische Übernahme der Wurzeldidaktik: Zunächst müssen Curriculum [FKR21], Lerngruppe [DDS15] und Lehr-Lernmethoden ausgearbeitet werden. Letztere werden vorwiegend aus den anderen Fächern herangezogen. Während Prinzipien der Datensammlung, -aufbereitung und -interpretation traditionell in den Sozialwissenschaften vermittelt werden [SHE98], werden Methoden zur Aussagekraft von Daten werden [KK01] aus der Statistik [Ha21] und Programmieren Lernen aus der Informatik [Sc11] hergeleitet. In diesem Rahmen ist eine genauere Kartierung der fachdidaktischen Einflüsse nicht umsetzbar. Stattdessen präsentiert diese Arbeit im Sinne der praxisnahen Entwicklungsforschung einen Beitrag zur fachdidaktischen Fortentwicklung, indem sie ein bereits erstelltes Curriculum, Lernprozessmodellierung und zugehöriges Kompetenzmodell vorstellt und konzeptionell mit Lerntechnologien integriert.

3 Lehr-Lernkonzept zur Beurteilung von Datenqualität

Zur Evaluation innovativer pädagogischer Praxis werden in Leuchtturmprojekten (best practice Beispielen), Alternativen zur statistischer Effektmessung der Lernergebnisse empfohlen [DWL17]. Denn neuartige Lehr-Lernpraxis kann nicht immer mit ausreichend hohen Teilnehmerzahlen durchgeführt werden, um die hohe Varianz von Lerntypen und kontextuellen Unterschieden auszugleichen. Weiterhin können motivationale langfristige Effekte, die beispielsweise durch situiertes Lernen erzeugt werden, nur selten mit kurzfristigen Leistungstests erfasst werden. Schlussendlich gibt es insbesondere bei computer-gestützten Lehr-Lernsettings einen verzerrenden Neuartigkeitseffekt, was bedeutet, dass das Novum der neuen Lernsituation schon ausreicht, um einen signifikanten Effekt zu begründen. Bei einer zweiten oder dritten Durchführung der Lerneinheit lässt dieser Effekt rapide nach. Als Alternativen zur klassischen Interventionsstudie gibt es zum Beispiel den Design-Based Research und Action Research. Im KODAQS Projekt wurde eine Form des letzteren Ansatzes implementiert. Aktionsforschung wird, aufgrund ihrer reflektierenden und iterativen Natur als eine sehr effektive Methode zur Bewertung und Verbesserung pädagogischer Praktiken geschätzt. Sie kombiniert theoretische Erkenntnisse mit praktischen Anwendungen und ermöglicht es Pädagogen, ihre Lehrstrategien systematisch zu verfeinern und sich an reale Bildungsumgebungen anzupassen [Le14; Mc13; SC21]. In diesem Sinne werden die Möglichkeiten und Grenzen des Lehr-Lernkonzepts von KODAQs vorgestellt, damit dieses schnell in der Praxis angewandt und verbreitet werden kann. Zudem werden erste Erfahrungsergebnisse und Best Practices, wie auch Desiderata für die weitere Entwicklung vorgestellt.

3.1 Bestehende Verfahren und Grenzen der Standardisierung von Datenqualität

Der Stand der Standardisierung oder auch Kanonisierung eines Faches bestimmt, wie offen der Lernprozess gestaltet werden kann. Hier gibt es ein Spannungsverhältnis zwischen der Selbststeuerung des Lernprozesses und der Vermittlung des in dem Fach als Kanon erachteten Anwendungswissens. Denn der autonome Lernprozess führt nicht notwendigerweise an den kanonischen Inhalten vorbei. Daher wird im Folgenden die inhaltliche Standardisierung des Wissens zu Datenqualität anhand einer aktuellen Überblicksstudie vorgestellt, um KODAQs in dem besagten Spannungsverhältnis zu verorten.

Daikeler et al. [Da24] geben einen systematischen Überblick über Datenqualitätskonzepte für moderne sozialwissenschaftliche Daten. Die Autor*innen haben einen Entscheidungsbaum für Forschende zur Identifikation geeigneter Datenqualitätskonzepte entwickelt. In dem entwickelten Entscheidungsbaum werden interne und externe Konzepte unterschieden. Interne Konzepte basieren auf den dateneigenen Fehlerquellen und Qualitätskriterien, die daher auch automatisiert bewertet werden können. Die Berechnung der zugehörigen Qualitätsindikatoren benötigt vertiefte Kenntnisse in der Statistik oder Informatik. Für die Prozessmodellierung bedeutet dies, dass die Lernenden standardisiertes Feedback zu

ihrer Datenqualität erhalten können, wenn sie ihre Daten im geeigneten Format angeben, da eine automatisierte Untersuchung von Datensätzen nur für bestimmte Dateiformate unterstützt werden kann. Externe Konzepte beziehen sich auf Gütekriterien, die nicht datenimmanent bestehen, sondern sich auf den Forschungsprozess, die Datenverfügbarkeit oder andere externe Qualitätskriterien beziehen. Hier wird eher sozialwissenschaftliche Expertise angesprochen. Neben der nur teilweise möglichen Automatisierung des Qualitätsfeedbacks ist auch die automatische Datenklassifikation, welche die Voraussetzung für automatisierte Qualitätsbewertung ist, noch in den Grundzügen. Dies wäre für den Aufbau einer Selbstlernumgebung notwendig, um entsprechende Algorithmen fallabhängig auszuführen. Diese Grenzen der technischen Entwicklung haben Auswirkungen auf die möglichen technologie-gestützten Ansätze, aber geben auch gewisse Prozessmodelle vor: Problem-basiertes oder selbstgesteuertes Lernen ist nur bei hoher Selbstkompetenz und fachlicher Expertise möglich. Damit verschiebt sich das Lehr-Lernmodell von der Nachfrageorientierung zur Angebotsorientierung. Dies bedeutet, dass der Lernprozess von den zu vermittelnden Inhalten und Kompetenzen der Lehrenden aus strukturiert wird, statt von den kognitiven oder entwicklungspsychologischen Schritten der Lernenden auszugehen.

3.2 Lehr-Lernkonzept der Data Quality Academy

Die Data Quality Academy hat zum Ziel, ein innovatives und bislang einzigartiges Lernangebot für die Sozialwissenschaften zu schaffen. Dieses Angebot richtet sich an Forschende, die ihre datenwissenschaftlichen Kompetenzen mit dem Schwerpunkt Datenqualität und qualitätsorientierte Nutzung von Daten ausbauen möchten und bietet flankierende Train-the-Trainer Angebote für die Vermittlung der Lehrinhalte in die Hochschullehre. Die Academy vermittelt als Lernangebot den Teilnehmenden die nötigen theoretischen und vor allem praktischen datenwissenschaftlichen Fähigkeiten, um die Qualität verschiedener Datenarten zu beurteilen und diese Daten qualitätsorientiert zur verlässlichen Beantwortung ihrer Forschungsfragen zu nutzen. Die Academy umfasst Lernangebote, die für verschiedene Zielgruppen mit ihren jeweiligen durch die Zielgruppenbefragung identifizierten Bedarfen und zeitlichen Verpflichtungen ausdifferenziert wurden. Damit können die Lernenden in Hackathon-ähnlichen Events, sogenannten „DataFests“, das erlernte direkt an eigener Forschung erproben.

Abbildung 2 und 3 enthalten eine Übersicht über die Kursplanung. Wie im letzten Kapitel begründet, ist diese angebotsorientiert und anwendungsorientiert. Im Gegensatz zu schulischen oder hochschuldidaktischen Lehr-Lernangeboten gibt es im Rahmen der Weiterbildung Forschender keine Notwendigkeit (z. B. wegen Selektions- oder Disziplinarzwängen) die erwarteten Lernerfolge kompetenzorientiert zu prüfen. Die Reihenfolge der Kursinhalte folgt daher eher dem Forschungsprozess als den Kompetenzanforderungen.

1	Theoretische und methodische Grundlagen		
Modul	**1.1** Konzepte und Rahmenwerke	**1.2** Indikatoren und Metriken	**1.3** Abhilfen und Korrekturen
Themen und Inhalte	• Was ist Datenqualität und welche Aspekte umfasst sie? • Wie manifestiert sich Datenqualität in verschiedenen Datentypen? • Welche speziellen ethischen und rechtlichen Aspekte sind zu beachten?	• Wie lässt sich Datenqualität diagnostizieren und quantifizieren? • Welche Indikatoren eignen sich für welchen Datentyp? • Wie beurteilt man die Ergebnisse?	• Wie lassen sich mögliche Probleme der Datenqualität korrigieren oder zumindest mildern? • Wie kann mit nicht korrigierbaren Qualitätsproblemen umgegangen werden?
Beispiele für Kurse	• „Datenqualität: Konzepte für digitale Verhaltensdaten“ • „Grundlagen der Erhebungen und Datenwissenschaft“ • „Kombination von Umfragedaten und digitalen Verhaltensdaten”	• Datenqualitätsindikatoren für Survey-Daten • Datenqualitätsindikatoren für digitale Verhaltensdaten • Analyse der Verzerrung durch Nichtbeantwortung	• „ Schritt für Schritt in der Umfragegewichtung “ • Multiple Imputation - Warum und wie • „Kalibrierung, Anpassungsgewichtung und Imputation“

Abb. 2: Teil 1 des Curriculums zur Vermittlung von Datenqualitätsmessungsverfahren

2	Praxis der qualitätsorientierten Arbeit mit Daten		
Modul	**2.1** Werkzeuge und Arbeitsabläufe	**2.2** Exkursionen (“Safaris”)	**2.3** Abschluss- und Zertifizierungsprojekt
Themen und Inhalte	• Forschungsdaten Administration • Daten Verknüpfung • Transparente, reproduzierbare Arbeitsabläufe • Wahlpflichtkurse zu speziellen Datentypen und Themen	• Wie wird Datenqualität in der Praxis gewährleistet und überprüft? → Exkursion zu einer Institution, die sozialwissenschaftliche Daten produziert und/oder bereitstellt • Reflexion & Transfer auf die eigene Arbeit mit Daten	• Wahlweise Teilnahme an einem „DataFest“ oder selbständige Bearbeitung eines Projekts aus dem eigenen Arbeitsbereich • Abgabe als reproduzierbares Notebook zur Nachnutzung durch Dritte

Beispiele für Kurse	• „Einführung in das Datenmanagement in der realen Welt“
	• „Moderne Arbeitsabläufe in der Datenwissenschaft„
	• „Verknüpfung von Twitter und Umfragedaten"
		• z.B.
		  • IAB
		  • DIW/SOEP
		  • Statistisches Bundesamt
		  • Wikimedia Foundation
		  • …
			• [selbst gewählter Datenschwerpunkt]
			• [selbst gewählter inhaltlicher Themenschwerpunkt]

Abb. 3: Teil 2 des Curriculums zur Vermittlung von Datenqualitätsmessungsverfahren

In der pädagogischen Forschung besteht Konsens, dass Lernprozesse besonders effektiv sind, wenn sie verschiedene Ebenen der kognitiven Anforderungen kombinieren. Zur Systematisierung wurde die Bloom'schen Taxonomie [AK01] entwickelt, die sich als Standardwerk etabliert hat. Basierend auf dieser lässt sich feststellen, dass eine vollständige Abdeckung der kognitiven Kategorien erreicht wird:

1. Erinnern: Konzepte und Rahmenwerke reproduzieren (1.1)
2. Verstehen: Indikatoren und Metriken inhaltlich durchdringen (1.2), Probleme in Daten korrigieren oder Abhilfe bei strukturellem Problem schaffen (1.3)
3. Anwenden: Anwendung der Konzepte auf das Forschungsdatenmanagement, Data Linkage (2.1)
4. Analysieren: Datenqualität in der Praxis analysieren (2.2), eigene Daten analysieren (2.1)
5. Bewerten: Transparenz und Reproduzierbarkeit bewerten (2.1), eigene Daten reflektieren und Wissen anwenden (2.2)
6. Erschaffen: Qualitativ hochwertiges Notebook und zugehörige Daten erschaffen und mittels Gütekriterien die Qualität der Daten reflektieren (2.3)

Gegeben die Zielgruppe der Forschenden, wie auch die Verquickung von Forschung und Lehre, lassen sich zwei Literaturtraditionen direkt zuordnen: Zum einen das forschende Lernen [HR19; Re12; Re20] und zum anderen der Scholarship of Teaching and Learning (SoTL) [Fe13; SH21]. Forschendes Lernen beschreibt die pädagogische Idee, den Lernprozess dem Forschungsprozess anzugleichen, wovon sich Motivation, Selbstwirksamkeit und Förderung von Nachwuchs erhofft wird. SoTL beschreibt eine Haltung, dass Lehre gleichzeitig mit Forschung verbunden wird. Ersteres lässt sich mit dem Motto „Lernen im Modus der Forschung", zweiteres mit „Lehre im Modus der Forschung" vereinfachen. Beide Konzepte passen für die Vermittlung von Datenqualität im Rahmen des KODAQS-Projektes. Denn die Anwendung von Datenqualitätsinhalten lässt sich direkt mit der eigenen Forschung verbinden, wodurch Sinnhaftigkeit und Motivation gestiftet werden soll. Ergänzend zum DataFest bieten auch die Gastaufenthalte bei KODAQS einen Raum zur begleiteten Forschung für externe Wissenschaftler*innen. Ferner sollen Teilnehmende der Academy – auch außerhalb der Teilnahme am DataFest – das Erlernte bewusst und kontrolliert auf ihre eigene Forschung anwenden.

Da dies kognitive komplexe Prozesse sind, die mit viel Unsicherheit verbunden sind, ergibt es Sinn, diese mit Reflexions- und Peer-Assessment zu verbinden [De21]. Reflexionsunterstützung in der Pädagogik bedeutet, Lernende dabei zu unterstützen, ihr eigenes Denken und Lernen zu analysieren und kritisch zu bewerten. Peer-Assessment bezieht sich auf die Methode, bei der Lernende die Arbeiten ihrer Mitschüler bewerten und konstruktives Feedback geben, was sowohl die eigene als auch die gegenseitige Lernentwicklung fördert. Durch diese beiden Methoden können Unsicherheiten bei der Selbststeuerung reduziert und tiefere Lernprozesse angeregt werden.

Befragungsergebnisse

Im Rahmen unserer Forschung zur Verbesserung akademischer Weiterbildungsangebote wurde ein mehrstufiges Untersuchungsdesign implementiert. Dazu gehörten ein Online-Fragebogen, Fokusgruppengespräche sowie ein Zielgruppenworkshop. Ziel war es, eine umfassende Rückmeldung zur Academy und deren Angebote zu erhalten und basierend darauf Verbesserungspotenziale zu identifizieren. Im Sinne der Aktionsforschung werden damit zum einen die vergangenen Weiterbildungsprinzipien der Lehrorganisation evaluiert als auch das Konzept der geplanten Academy validiert.

1. Online-Fragebogen: Wir haben einen ca. 10-minütigen Online-Fragebogen mit sozialwissenschaftlich Forschenden geteilt. Von den 18 Teilnehmenden waren 50% weiblich, 38% männlich und 12% divers oder ohne Angabe. Die Teilnehmenden setzten sich aus 50% Doktorand*innen und 50% Postdocs zusammen. Fachlich waren 50% aus der Soziologie, 25% aus der Politikwissenschaft, 19% aus der Psychologie und 6% aus anderen Disziplinen vertreten.
2. Fokusgruppengespräche: Wir führten zwei einstündige Fokusgruppengespräche mit je zwei Professor*innen der Sozialwissenschaften, Psychologie und sozialwissenschaftlichen Methodenlehre durch.
3. Zielgruppenworkshop: Zusätzlich organisierten wir einen 1,5-stündigen Workshop mit sieben (Post-)Doktorand*innen der Sozialwissenschaften und Psychologie.

Die Ergebnisse zeigten eine große inhaltliche Diversität, sowohl innerhalb als auch zwischen den Disziplinen: Die Teilnehmenden schätzten unterschiedliche Themen als variierend relevant ein, was auf die Sinnhaftigkeit der modularen Kursstruktur hinweist. Die Befragten betonten weiterhin die Wichtigkeit eines bedarfsorientierten Ansatzes und eines praxisnahen Umgangs mit Datenqualitätsproblemen

Es wurden spezifische Themen wie der Umgang mit Datenqualitätsproblemen, amtliche Daten und ihre Tücken, Gewichtung und Stichprobentheorie sowie die Auswirkungen von Datenqualitätsaspekten auf inhaltliche Schlussfolgerungen hervorgehoben.

Die Rückmeldungen zum Format der Academy betonten:

- Lernen voneinander: Präsenzveranstaltungen wurden als wichtig erachtet, um vertiefende Fragen zu klären und voneinander zu lernen.
- Homogenität der Gruppe: Eine fachliche Homogenität wurde als hilfreich betrachtet, während eine homogene Seniorität als weniger wichtig angesehen wurde.
- Struktur der Online-Formate: Es wurde vorgeschlagen, dass Online-Formate einem einheitlichen Schema folgen sollten (z.B. immer 30 Minuten).
- Arbeitnehmer*innenfreundlichkeit: Es wurde betont, dass die Formate arbeitnehmer*innenfreundlich gestaltet sein sollten.

Die Teilnehmenden äußerten verschiedene Motivationen für die Teilnahme an dem recht zeitintensiven Curriculum. Ein zentraler Aspekt war die bedarfsorientierte Begleitung, wobei die Inhalte die Arbeit an einem aktuellen Projekt unterstützen sollten. Auch interdisziplinäre Projekte wurden als motivierend betrachtet; ein gruppenbasiertes, interdisziplinäres Publikationsprojekt könnte die Motivation deutlich steigern. Darüber hinaus sollten die Inhalte und Materialien direkt in der Lehre verwendet werden können, insbesondere neue Daten und deren didaktische Umsetzung. Die Möglichkeit, sich mit anderen Forschenden zu vernetzen und Probleme gemeinsam zu diskutieren, wurde ebenfalls als wertvoll erachtet.

Basierend auf den erhobenen Daten lassen sich folgende Handlungsempfehlungen ableiten: Eine modulare Struktur der Kurse mit individuellen Auswahlmöglichkeiten würde der inhaltlichen Diversität gerecht werden. Der Fokus der Academy sollte auf der praktischen Anwendung liegen, wobei der Anwendungsbezug auch in den theoretischen Inhalten erkennbar sein sollte. Präsenzveranstaltungen sollten einen festen Bestandteil des Curriculums bilden, wobei die Gruppen fachlich homogen sein sollten. Online-Formate sollten einer einheitlichen Struktur folgen und arbeitnehmer*innenfreundlich gestaltet sein. Die Academy eignet sich als Train-the-Trainer-Konzept, um die Anwendung der Inhalte in der Lehre zu fördern. Schließlich besteht hohes Interesse und Zustimmung zur praktischen Anwendung der Inhalte durch die Arbeit an eigenen oder vorgegebenen Projekten.

Desiderata und weiterführende Forschung

Der Erfahrungsbericht wurde in einer frühen Phase des Projektes erstellt. Das KODAQS Projekt zeigt vielversprechende innovative Ideen zur Weiterentwicklung der Didaktik der Datenqualitätsmessung, wobei es direkt in der Praxis und Forschung verwurzelt ist. Weitere praxisnahe Publikationen von Best Practices und begleitender Forschung sind geplant.

Nähere Informationen zu den Lernerfolgen und dem motivationalen und konzeptionellen Erfolg stehen noch aus. Dennoch zeichnet sich ein Bild ab, dass die Teilnehmenden gruppenorientiert, projektbasiert und vor allem im Modus der Forschung lernen wollen. Dies ist vor allem in den späteren Modulen des Curriculums udn insbesondere im Abschlussmodul "DataFest"verortet. Forschendes Lernen von Anfang an würde das Projekt vor organisatorische und konzeptionelle Probleme stellen. Denn die Vielfalt der möglichen Interessen und Lernpfade kann in Live-Workshops kaum antizipiert und vorbereitet werden.

Computergestütztes Lernen könnte Abhilfe für dieses Problem leisten, indem autonomere Lernmethoden ermöglicht werden. Wünschenswert wäre hierfür:

1. Systematisierung der Qualitätsindikatoren für die relevanten Datentypen
2. Ein technischer Standard, der Datentyp, mögliche Qualitätsprobleme, automatische und semi-automatische Bewertung der Datenqualität beschreibt
3. Eine E-Learning-Umgebung, bei der Lernende ihre Daten mitbringen können, und semi-automatisierte Bewertung, aber auch passende Lernressourcen abrufen können
4. Bereitstellen der Lernmodule als Open Educational Resources (OER)

Eine höhere Projekt- und Gruppenorientierung wurde sowohl in der pädagogischen Modellierung als auch in den Konzeptevaluierung angeraten. Hierfür lassen sich folgende Desiderata formulieren:

1. Forschendes Lernen (Durchbrechen des Inhalt-vor-Anwendung Paradigmas) in der Academy ausbauen
2. Literaturbasierte Kartierung der Fachdidaktik von Data Quality in angrenzenden Gebieten erweitern
3. Fachdidaktik aktionsbasiert weiterentwickeln, so dass den Lerninhalten Lernformen (Gruppenarbeit, Projektarbeit, Selbstlernen etc.) und Lernpfade wie auch ausformulierte Kompetenzen zugeordnet werden können, indem ein Kompetenzmodell für Datenqualitätsmessung erstellt wird
4. Evidenzbasierte Begleitforschung im Sinne des Scholarship of Teaching and Learning (SoTL) durchführen

Literaturverzeichnis

[ABK20] Amaya, A.; Biemer, P. P.; Kinyon, D.: Total Error in a Big Data World: Adapting the TSE Framework to Big Data. Journal of Survey Statistics and Methodology 8 (1), S. 89–119, 2020, DOI: 10.1093/jssam/smz056, URL: https://academic.oup.com/jssam/article/8/1/89/5728725, Stand: 19. 12. 2023.

[AK01] Anderson, L. W.; Krathwohl, D. R.: A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: complete edition. Addison Wesley Longman, Inc., 2001.

[BM23] BMBF: Nationale Forschungsdateninfrastruktur, de, 2023, URL: https://www.bmbf.de/bmbf/de/forschung/das-wissenschaftssystem/nationale-forschungsdateninfrastruktur/nationale-forschungsdateninfrastruktur_node.html, Stand: 10. 05. 2024.

[Br23] Brodie, M. L.: Defining data science: a new field of inquiry. Publisher: arXiv, 2023, URL: https://arxiv.org/abs/2306.16177.

[Da24] Daikeler, J. et al.: Assessing Data Quality in the Age of Digital Social Research: A Systematic Review. en, Social Science Computer Review, S. 08944393241245395, 2024, DOI: 10.1177/08944393241245395, Stand: 08. 05. 2024.

[DDS15] D. Asamoah; Derek Doran; Shu Z. Schiller: Teaching the Foundations of Data Science: An Interdisciplinary Approach. arXiv.org, 2015, URL: https://arxiv.org/abs/1512.04456.

[De21] Dehne, J.: Möglichkeiten und Limitationen der medialen Unterstützung forschenden Lernens, deu, Diss., Universität Potsdam, 2021, DOI: 10.25932/publishup-49789, URL: https://publishup.uni-potsdam.de/frontdoor/index/index/docId/49789, Stand: 09.09.2023.

[DWL17] Dehne, J.; Wiepke, A.; Lucke, U.: Evaluierung von E-Learning - Ein Kommentar zu "Media will Never Influence Learning". In (Igel, C.; Ullrich, C.; Wessner, M., Hrsg.): Bildungsräume 2017: DeLFI 2017, Die 15. e-Learning Fachtagung Informatik, der Gesellschaft für Informatik e.V. (GI), 5. bis 8. September 2017, Chemnitz. Bd. P-273. LNI, Gesellschaft für Informatik, Bonn, S. 167–177, 2017, URL: https://dl.gi.de/20.500.12116/4838.

[FBD23] Fröhling, L.; Birkenmaier, L.; Daikeler, J.: Garbage in-Garbage out? Datenqualität im Umgang mit digitalen Verhaltensdaten. easy_social_sciences (68), S. 21–30, 2023.

[Fe13] Felten, P.: Principles of Good Practice in SoTL. Teaching & Learning Inquiry: The ISSOTL Journal 1 (1), Publisher: [Indiana University Press, International Society for the Scholarship of Teaching and Learning], S. 121–125, 2013, DOI: 10.2979/teachlearninqu.1.1.121, URL: https://www.jstor.org/stable/10.2979/teachlearninqu.1.1.121, Stand: 28.05.2024.

[FKR21] Fekete, A.; Kay, J.; Rohm, U.: A Data-centric Computing Curriculum for a Data Science Major. In: Proceedings of the 52nd ACM Technical Symposium on Computer Science Education. ACM, Virtual Event USA, 2021.

[GL10] Groves, R.M.; Lyberg, L.: Total Survey Error: Past, Present, and Future. en, Public Opinion Quarterly 74 (5), Publisher: Oxford University Press, S. 849–879, 2010, DOI: 10.1093/poq/nfq065, URL: https://academic.oup.com/poq/article-lookup/doi/10.1093/poq/nfq065, Stand: 19.12.2023.

[Ha21] Haensch, A.-C. et al.: The International Program in Survey and Data Science (IPSDS): A modern study program for working professionals. en, Statistical Journal of the IAOS 37 (3), S. 921–933, 2021, DOI: 10.3233/SJI-210833, URL: https://www.medra.org/servlet/aliasResolver?alias=iospress&doi=10.3233/SJI-210833, Stand: 03.05.2024.

[HR19] Huber, L.; Reinmann, G.: Forschungsnahes Lernen verstehen: Begriff und Genese. In (Huber, L.; Reinmann, G., Hrsg.): Vom forschungsnahen zum forschenden Lernen an Hochschulen: Wege der Bildung durch Wissenschaft. Springer Fachmedien, Wiesbaden, S. 1–28, 2019, DOI: 10.1007/978-3-658-24949-6_1, URL: https://doi.org/10.1007/978-3-658-24949-6_1, Stand: 28.05.2024.

[HSW07] Herzog, T.N.; Scheuren, F.; Winkler, W.E.: Data quality and record linkage techniques. OCLC: ocn137313060, Springer, New York ; London, 2007.

[IS20] ISO: Data quality, ISO 8000-2:2020, Geneva, Switzerland: International Organization for Standardization, 2020, URL: https://www.iso.org/standard/80543.html.

[KBP15] Kyng, T.; Bilgin, A.; Puang-Ngern, B.: Data science: is it statistics or computer science? Statistics education in the age of big data. In: Advances in Statistics Education: Developments, Experiences, and Assessments IASE Satellite. International Association for Statistical Education, 2015.

[KK01] Kühnel, S.-M.; Krebs, D.: Statistik für die Sozialwissenschaften. Grundlagen, Methoden, Anwendungen. Reinbek bei Hamburg: Rowohlt, 2001.

[Le14] Lesha, J.: Action Research In Education. European Scientific Journal 10 (13), 2014.

[Lö19] Löffler, F. et al.: RSE4NFDI - Safeguarding software sustainability in the NFDI, en, 2019, DOI: 10.5281/zenodo.2630451.

[Mc13] McAteer, M.: Action Research in Education. Sage/BERA, London, 2013.

[Re12] Reinmann, G.: Studientext Evaluation, Place: München, 2012, Stand: 14. 05. 2019.

[Re20] Reinmann, G.: Forschungsnahes Lehren und Lernen an Hochschulen in der Denkfigur des didaktischen Dreiecks. In (Brinkmann, M., Hrsg.): Forschendes Lernen: Pädagogische Studien zur Konjunktur eines hochschuldidaktischen Konzepts. Springer Fachmedien, Wiesbaden, S. 39–59, 2020, DOI: 10.1007/978-3-658-28173-1_3, URL: https://doi.org/10.1007/978-3-658-28173-1_3, Stand: 28. 05. 2024.

[Sc11] Schubert, S. et al.: Didaktik der Informatik. Springer, 2011.

[SC21] Saez Bondia, M.; Cortes Gracia, A.: Action research in education: a set of case studies? Educational Action Research 30 (5), Publisher: Informa UK Limited, S. 850–864, 2021.

[Se21] Sen, I. et al.: A Total Error Framework for Digital Traces of Human Behavior on Online Platforms. Public Opinion Quarterly 85 (S1), S. 399–422, 2021, DOI: 10.1093/poq/nfab018, URL: https://academic.oup.com/poq/article/85/S1/399/6359490, Stand: 19. 12. 2023.

[SH21] Steiner, H. H.; Hakala, C. M.: What Do SoTL Practitioners Need to Know about Learning? Teaching and Learning Inquiry 9 (1), Number: 1, S. 79–85, 2021, DOI: 10.20343/teachlearninqu.9.1.7, URL: https://journalhosting.ucalgary.ca/index.php/TLI/article/view/70878, Stand: 28. 05. 2024.

[SHE98] Schnell, R.; Hill, P. B.; Esser, E.: Methoden der empirischen Sozialforschung. Oldenbourg Wissenschaftsverlag, München Wien, 1998.