

Qualität und Nachnutzungspotenzial von LIS-Forschungsdaten in offenen Repositorien

Katharina Schobert¹, Tamara Heck², Marc Rittberger^{1,2}

¹ DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation,
Frankfurt am Main, Deutschland

² Hochschule Darmstadt, Deutschland

{[k.schobert](mailto:k.schobert@difp.de), [t.heck](mailto:t.heck@difp.de), [m.rittberger](mailto:m.rittberger@difp.de)}@difp.de

Abstract

Die Bereitstellung von Forschungsdaten kann in den meisten Disziplinen noch nicht als Standard bezeichnet werden, aber immer mehr Forschende publizieren ihre Daten in offenen Repositorien. Hierbei sind es meist die Ersteller*innen der Daten, die diese mit Metadaten versehen und beschreiben. Neben der eigentlichen Ergebnispublikation erfordert die ordentliche Dokumentation und Beschreibung der Daten nochmals Kompetenzen und Kapazitäten, damit die Daten von Dritten auch verstanden und potenziell nachgenutzt werden können. Darüber hinaus sind die Vielzahl der Datenrepositorien groß und die Portale sehr heterogen bezüglich Möglichkeiten zur Beschreibung von Datensätzen. Im Rahmen des Forschungsdatenmanagements existieren Standards zur Beschreibung von Datensätzen. Jedoch besteht hierbei die Herausforderung, dass die Standards die unterschiedlichen Methoden der Datenerhebung und -analyse ausreichend darstellen. Diese Work-in-Progress-Studie hat verschiedene offene Repositorien auf Datensätze aus der Informationswissenschaft durchsucht und diese auf ihre Qualität und ihr Nachnutzungspotenzial hin überprüft. Als Grundlage hat diese Studie Qualitätskriterien eines Forschungsdatenzentrums aus der Bildungsforschung sowie die FAIR-Prinzipien genutzt. Die Studie geht der Frage nach, welche LIS-Daten publiziert und beschrieben werden und ob die verwendeten Qualitätskriterien auf diese Datensätze gut anwendbar sind. Ausgewertet wurden 44 Datensätze aus sechs fachübergreifenden offenen Repositorien. Die Auswertungen zeigen, dass die meisten LIS-Datensätze die festgelegten Kriterien gut erfüllen, was ihre Verständlichkeit und potenzielle Nachnutzbarkeit unterstützt. Herausforderungen gibt es jedoch bei der Anwendung der vorhandenen Qualitätskriterien, die Anpassungen und methodenspezifische Ansätze erfordern. Zusätzlich wurden Schwierigkeiten bei der Datensatzrecherche identifiziert. Die Ergebnisse der

Studie geben Hinweise darauf, wie die Auffindbarkeit und somit die Nachnutzung von LIS-Datensätzen verbessert werden kann.

Keywords: Forschungsdaten, Repositorium, Forschungsdatenmanagement, Informationswissenschaft, Open Science

1 Einleitung

In den letzten Jahren ist die Menge an Forschungsdaten, die in öffentlichen Repositorien zur Verfügung stehen, stark gewachsen. Diese wachsende Menge an Forschungsdaten eröffnet der wissenschaftlichen Gemeinschaft enorme Potenziale (Bambey et al., 2017). Der Zugriff auf die umfangreichen Datenbestände erlaubt es Wissenschaftler*innen, publizierte Forschungsergebnisse besser nachzuvollziehen, zu überprüfen und die Daten nachzunutzen, um entweder eigene erhobene Daten anzureichern oder mit den schon vorhandenen Daten eigene Forschungsfragen zu beantworten (ebd.; Rahal et al., 2022). Letzteres kann eine enorme Zeitersparnis für Forschende bedeuten, da sie eigene Daten für ihre Untersuchung nicht erst erheben müssen (Tenopir et al., 2020).

Um die Archivierung und den Zugang zu qualitativen Forschungsdaten zu erleichtern, hat der RatSWD¹ Empfehlungen zur Etablierung von Standards für die Archivierung und den Zugang zu Forschungsdaten, zur Entwicklung von Metadatenstandards und zur Schaffung von Repositorien für Forschungsdaten herausgegeben (Bambey/Corti et al., 2018). Damit Forschungsdaten später effektiv genutzt werden, ist es von großer Bedeutung, dass sie zuerst gefunden werden können (Fecher/Puschmann, 2015; Wallis et al., 2013). Hierbei spielen öffentliche Repositorien wie Universitäts-Repositorien und Forschungsdatenzentren (FDZ) eine entscheidende Rolle. Plattformen wie re3data agieren darüber hinaus als Metasuche und erlauben eine Recherche nach verfügbaren Datenquellen (Pampel et al., 2013; Watteler, 2022). Durch diese Infrastrukturen soll die gezielte Suche nach Forschungsdaten erleichtert und eine effiziente Nachnutzung ermöglicht werden (Feldsien-Sudhaus/Rajski, 2016; Pampel et al., 2013).

1 <https://www.konsortswd.de/ratswd>

Trotz der Potenziale birgt die Nutzung von öffentlichen Datenrepositorien Herausforderungen hinsichtlich der Qualität, Relevanz und Nachnutzung der Forschungsdaten (Neuroth/Oevel, 2021). Eine zentrale Herausforderung besteht in der Beurteilung der Qualität von Forschungsdaten, wobei hier unterschiedliche Faktoren relevant sind (Bishop/Kuula-Luumi, 2017; Fecher/Puschmann, 2015; Wallis et al., 2013). Faktoren für die Qualität von Forschungsdaten sind bspw. die Verständlichkeit der Beschreibung und Bereitstellung der Daten, aber auch die Nützlichkeit und Nutzbarkeit hinsichtlich der eigenen Forschung (Bambey/Meyermann et al., 2018). Des Weiteren spielt die Qualität der Studie, in der die Daten erhoben wurden, eine Rolle (Bambey/Meyermann et al., 2018; Kindling, 2013). Aufgrund der Vielfalt der Datenquellen, Formate und Datenerfassungsmethoden ist es schwierig, eine einheitliche Qualitätssicherung zu gewährleisten (Kindling, 2013). Es bedarf jedoch geeigneter Kriterien, um die Datenqualität zu bewerten, sodass Forschende Daten nachnutzen können (Bishop/Kuula-Luumi, 2017; Kindling, 2013; Neuroth/Oevel, 2021).

Damit Forschende publizierte Datensätze besser finden und hinsichtlich ihrer Nachnutzbarkeit bewerten können, legen Betreiber von Forschungsdatenzentren und Repositorien Standards für Metadaten zur Beschreibung der Daten fest. Diese basieren häufig auf Leitlinien von Konsortien wie der Nationalen Forschungsdateninfrastruktur² oder dem Rat für Sozial- und Wirtschaftsdaten in Deutschland (RatSWD). In Forschungsdatenzentren werden Datensätze und ihre Beschreibungen in der Regel vor der Veröffentlichung eines Datensatzes von Kurator*innen abgefragt, geprüft und bei Bedarf angereichert. In offenen Online-Repositorien sind die Forschenden meist selbst dafür verantwortlich, ihre Datensätze so zu beschreiben, dass sie für Dritte verständlich und auffindbar sind (Pampel et al., 2013). In diesem Fall müssen also die Forschenden selbst entscheiden, wie und mit welcher Dokumentation sie ihre Daten publizieren (Watteler, 2022). Andererseits sind sie von den Gegebenheiten der Repositorien und deren Bereitstellung von Metadatenfeldern abhängig. Beides kann sich auf die Qualität der zur Verfügung gestellten Daten und somit auf die Suche und Auffindbarkeit als auch auf die Nachnutzung von Daten auswirken (Kindling, 2013; Piwowar/Vision, 2013; Watteler, 2022).

Die Pilotstudie hat untersucht, welche Forschungsdaten aus dem Bereich Bibliotheks- und Informationswissenschaft (LIS) offen zugänglich sind und

2 <https://www.nfdi.de>

wie diese beschrieben werden. Sie geht den Forschungsfragen nach, ob LIS-Datensätze qualitativ so abgelegt sind, dass sie verständlich und potenziell gut nachnutzbar sind, und ob existierende Qualitätskriterien ausreichend sind, um die Daten bestmöglich zu beschreiben.

Abschnitt 2 führt die Qualitätskriterien ein und diskutiert weitere Herausforderungen bei der Bereitstellung und Nachnutzung von Forschungsdaten. Abschnitt 3 beschreibt die Methode der Datensammlung und Auswahl der Datensätze, bevor in Abschnitt 4 die Ergebnisse vorgestellt und diskutiert werden. Abschnitt 5 zieht das Fazit der Arbeit.

2 Forschungsdaten und Qualitätskriterien

Forschungsdaten sind Daten, die im Rahmen wissenschaftlicher Untersuchungen erhoben, generiert, verarbeitet und ausgewertet werden. Diese können in verschiedenen Formaten vorliegen, z.B. als Text-, Audio- oder Videodateien, als Messwerte, als Bilder oder als Softwarecode. Forschungsdaten können sowohl quantitative als auch qualitative Informationen enthalten und können aus verschiedenen Quellen stammen, z.B. aus Beobachtungen, Experimenten, Umfragen oder Simulationen, aber auch aus Datensammlungen aus dem Web oder historischen Beständen (Rösch, 2021). Forschungsdaten sind oft das Ergebnis aufwendiger und kostenintensiver Forschungsprozesse und stellen somit einen wertvollen Forschungsrohstoff dar, der in der Regel über einen längeren Zeitraum hinweg aufbewahrt und nachnutzbar gemacht werden sollte (Tenopir et al., 2020).

Forschungsinstrumente sind Werkzeuge, die im Rahmen wissenschaftlicher Untersuchungen eingesetzt werden, um Daten zu erheben, zu generieren, zu verarbeiten und auszuwerten. Diese können sowohl physischer als auch virtueller Natur sein und reichen von einfachen Fragebögen über Laborgeräte und Beobachtungssysteme bis hin zu komplexen Software-Tools für die Datenanalyse (Bambey/Rittberger, 2013). Forschungsinstrumente dienen dazu, Forschungsfragen zu beantworten oder Hypothesen zu überprüfen, indem sie empirische Daten sammeln und systematisch auswerten. Sie müssen in der Regel validiert und reliabel sein, d.h. sie müssen geeignet sein, um genau das zu messen oder zu erfassen, was sie messen oder erfassen sollen, und müssen zuverlässig und reproduzierbar sein, um valide Ergebnisse zu liefern (ebd.).

In Rechercheportalen von professionellen Forschungsdatenzentren kann oft nach Forschungsdaten und Forschungsinstrumenten gesucht werden. In offenen Repositorien finden sich unter „Datensätze“ häufig auch Instrumente, bspw. unterscheidet das Portal Zenodo³ nicht nach Daten und Instrumenten. Da beides benötigt wird, um wissenschaftliche Untersuchungen durchzuführen und empirische Daten zu sammeln und auszuwerten, liegt es auf der Hand, Instrumente und Daten als gemeinsame Dokumentationseinheit in einem Repository zu veröffentlichen.

Für die Recherche und die potenzielle Nachnutzung von Daten sind verschiedene Qualitätsaspekte relevant. Wenn Forschende Daten recherchieren und nachnutzen wollen, werden sie die gefundenen Ressourcen anhand spezifischer Merkmale bewerten (Bishop/Kuula-Luumi, 2017; Kindling, 2013). Ihr Ziel ist es, nutzbare und valide Daten für ihr Forschungsvorhaben zu finden. Studien zeigen, dass nicht alle Forschende mit der Suche nach geeigneten Daten in ihrer Disziplin zufrieden sind (Ackermann et al., o.J.; Lösch et al., 2022; Scherp et al., 2020).

Eine sorgfältige Dokumentation und detaillierte Beschreibung der publizierten Daten, die ihre Herkunft, Verarbeitungsschritte und potenzielle Einschränkungen transparent machen, trägt wesentlich zur Auffindbarkeit von Daten bei. Dies gilt insbesondere dann, wenn die Dokumentation und Beschreibung nach einheitlichen Regeln und Standards erfolgen (Fecher/Puschmann, 2015; Neuroth/Oevel, 2021; Neuroth et al., 2019). Forschungsdatenzentren wie das FDZ Bildung haben gemeinsam mit Partner*innen Leitlinien zum Datenmanagement entwickelt (Eckert et al., 2023): Ein standardisierter Datenmanagementplan (STAMP⁴), der speziell für Studien in der Bildungsforschung entwickelt wurde, bietet nicht nur Unterstützung beim Datenmanagement, sondern auch bei der Beschreibung der Daten im Einklang mit den FAIR-Prinzipien (Wilkinson et al., 2016). Darüber hinaus gibt STAMP Hinweise zur Dokumentation der dort häufig durchgeführten Methoden und Forschungsdesigns. Am FDZ eingehende Daten werden auf ihre Beschreibung und Transparenz hin überprüft und mit Metadaten angereichert, um sie besser auffindbar zu machen. Die Dokumentation der Daten und Instrumente hängt dabei wesentlich von der verwendeten Methode und dem Design ab. Denn unterschiedliche Methoden und Datenerhebungen erfordern spezifische oder teils ausführlichere Metadaten und Beschrei-

3 <https://zenodo.org>

4 <https://www.forschungsdaten-bildung.de/stamp-nutzen>

bungen, damit Dritte die Relevanz der Nachnutzung schnell ersichtlich beurteilen können. Eine Herausforderung beim Datenmanagement ist somit die Unterstützung durch Standards zur Dokumentation und Archivierung von diversen Daten und Instrumenten.

3 Methode

Im Folgenden werden die Generierung der Datensätze und die zur Untersuchung verwendeten Kriterien vorgestellt.

3.1 Recherche der Datensätze

Für die Datensammlung wurden zunächst verschiedene disziplinübergreifende öffentliche Repositorien identifiziert, die auf einschlägigen Webseiten⁵ und in der Literatur (Watteler, 2022) genannt werden: Dryad⁶, Figshare⁷ und Zenodo⁸. Dataverse wurde zunächst aufgenommen, jedoch aufgrund der großen Anzahl an Datensätzen, die mit den vorhandenen Filtern nicht effektiv auf das Fachgebiet eingegrenzt werden konnten, letztendlich ausgeschlossen. Darüber hinaus wurden Repositorien von Hochschulen durchsucht, deren LIS-Institute Mitglied im Hochschulverband Informationswissenschaft⁹ sind. Tabelle 1 zeigt die durchsuchten Quellen. Es ist anzumerken, dass nicht alle Institute (laut Angaben auf Website) eigene Datenrepositorien haben bzw. in manchen Fällen keine Repositorien gefunden wurden.

Die Recherche nach LIS-Datensätzen wurde an die Optionen in den entsprechenden gefundenen Repositorien angepasst. Um die Suche handhabbar zu machen, wurden alle Quellen aufgenommen, die entweder nur Datensätze enthalten oder bei denen es möglich ist, Ergebnisse nach Datensätzen zu filtern. Wo es möglich und sinnvoll war, wurden neben dem Filter „Datensatz“ auch der Filter für den Fachbereich genutzt. In einigen Fällen erfolgte die Suche direkt mit deutsch- und englischsprachigen Suchbegriffen in verschiedenen Schreibweisen, da eine entsprechende Filterung entweder nicht

5 z.B. <https://www.nature.com/sdata/policies/repositories#general>

6 <https://datadryad.org/>

7 <https://figshare.com>

8 <https://zenodo.org>

9 <https://www.informationswissenschaft.org/institutionen>

verfügbar oder nur eingeschränkt möglich war. In Dryad und Figshare wurde aufgrund der umfangreichen Datenmenge lediglich mit einem einzigen Suchbegriff gesucht.

Tab. 1: Auswahl der Repositorien für die Datensatzrecherche

HI-Mitglieder (Stand 4/2023)	gefundenes Repository	Hinweis	ID
HU Berlin	https://edoc.hu-berlin.de/		1
Hochschule Darmstadt	https://zenodo.org/communities/h_da/?page=1&size=20	kein eigenes Repository, Daten sollen über Zenodo geteilt werden laut Angabe	2
HHU Düsseldorf	https://researchdata.hhu.de		3
HAW Hamburg	https://reposit.haw-hamburg.de/		4
FH Hannover	https://serwiss.bib.hs-hannover.de/home	nur Schriften, keine Filterung nach Daten möglich	/
SU Hildesheim	https://hildok.bsz-bw.de		5
TH Köln	https://publiscologne.th-koeln.de/home		6
Uni Konstanz	https://kondata.uni-konstanz.de/		7
HTWK Leipzig	https://htwk-leipzig.qucosa.de	keine Filterung nach Daten möglich	/
Archivschule Marburg	https://archiv.ub.uni-marburg.de/ubfind/Browse/Home	E-Papers, keine Filterung nach Daten möglich	/
FH Potsdam	https://opus4.kobv.de/opus4-fhpotsdam/home		8
Uni Regensburg	https://epub.uni-regensburg.de		9
HdM Stuttgart	https://hdms.bsz-bw.de/home		10
FH Burgenland, Eisenstadt	https://www.fh-burgenland.at/studieren/bibliothek/digital-repository/	keine Angaben zu Forschungsdaten	/
KFU Graz		AUSSDA für SoWi, GAMS für Dig. Hum., kein fachübergreifendes Repository laut Angabe)	/
FH Graubünden, Chur		kein Repository oder Angaben gefunden	/

HI-Mitglieder (Stand 4/2023)	gefundenes Repository	Hinweis	ID
Uni Bern	https://boris-portal.unibe.ch/		11
FH Genf		kein eigenes Repository laut Angabe	/
Uni Zürich	https://www.zora.uzh.ch/	nur für Publikationen, haben kein eigenes Datenrepository laut Angabe	/
Disziplinübergreifende Repositorien			
Dryad	https://datadryad.org/stash		12
Figshare	https://figshare.com		13
Zenodo	https://zenodo.org		14

Die Recherchen in 14 Repositorien wurden von April bis Juni 2023 durchgeführt und führten zu 98 Treffern aus 14 verschiedenen Repositorien. Die gefundenen Treffer wurden dahingehend geprüft, ob diese dem LIS-Bereich zugeordnet werden können. Forschungsdaten aus Studien, die eindeutig nicht dem LIS-Bereich zugeordnet werden können, wurden ausgeschlossen – ebenso Ressourcen, bei denen es sich um Trainingsmaterialien (Präsentationsfolien), Handbücher oder reine Dokumentationen zu Studien handelte ($n=4$). Dies führte zu einer Menge von 44 Datensätzen aus sechs Repositorien, die auf Qualitätskriterien hin untersucht wurden (Schobert et al., 2023). Tabelle 2 zeigt die Rechercheanfragen per Repository.

Tab. 2: Recherche und Prüfung der Datensätze

ID	Suchanfrage oder Filter	Anzahl Treffer	Anzahl Treffer nach Prüfung
1	Filter: Forschungsdaten; DDC 000 und DDC 020	17	15
2	Einfache Suche*: "Information Science" OR "information science" OR "informationscience" OR "Informationswissenschaft" OR "informationswissenschaft"	0	0
	Hinweis: keine Datensets vorhanden		
3	Einfache Suche*	9	0
4	Filter: Datenset; Schlagwort: DDC 020 und 004	4	4
5	Filter: alle Dokumente; kein Dokumenttyp "Datensatz" vorhanden	0	0

ID	Suchanfrage oder Filter	Anzahl Treffer	Anzahl Treffer nach Prüfung
6	Einfache Suche*; kein Dokumenttyp "Datensatz" vorhanden	0	0
7	Filter: Klick auf "Daten durchsuchen"; Fachgebiet "Social Sciences" und "Environmental Science and Ecology"	4	0
8	Filter: Klick auf "alle Dokumente"; kein Dokumenttyp "Datensatz" vorhanden	0	0
9	Filter: DDC 020; Dokumentenart "Datensatz"	1	1
10	Klick auf "alle Dokumente"; kein Dokumenttyp "Datensatz" vorhanden	0	0
11	Filter: Datensets; DDC 500	18	0
12	Einfache Suche: "information science"	20	9
13	Filter: Typ: Datenset; Kategorie: "information science"	10	3
14	keywords:"Information Science" OR keywords:"information science" OR keywords:"informationscience" OR keywords:"Informationscience" OR keywords:"Informationswissenschaft" OR keywords:"informationswissenschaft"; Filter: Typ: Datenset; Rechte: offen	15	12
Summe		98	44

3.2 Qualitätskriterien

Die Qualitätskriterien wurden 2020 im Rahmen einer Masterarbeit entwickelt und basieren zum einen auf Qualitätskriterien des Forschungsdatenzentrums Bildung als auch auf den FAIR-Prinzipien (Bambey/Rittberger, 2013; Wilkinson et al., 2016). Die Masterarbeit untersuchte damals rund 40 Datensätze aus der Bildungsforschung (Heck et al., 2021; Heck/Schneider, 2021). Tabelle 3 zeigt die verwendeten Kriterien sowie deren Messung. Die Arbeit wurde von drei Personen in enger Abstimmung miteinander durchgeführt.

Tab. 3: Qualitätskriterien und deren Prüfung

Kriterien ID	Name	Beschreibung	Quelle	Prüfung
K1	Autor*in IDs	Ist Urheberschaft eindeutig identifizierbar (ID)?	FAIR F.1, R1.2	2={ORCID pro Autor*in}; 1=ORCID Hauptautor*in}; 0={nein}
K2	Ort Datenerhebung	Ist der Erhebungsort genannt?	FDZ, R1.2	2={ja, im Abstract}; 1={ja, in Datei}; 0={nein}
K3	Zeit Datenerhebung	Ist der Zeitpunkt der Erhebung genannt?	FDZ	2={ja, im Abstract}; 1={ja, in Datei}; 0={nein}
K4	Anzahl Objekte	Ist die Anzahl der beforschten Personen/ Objekte angegeben?	FDZ	2={ja, im Abstract}; 1={ja, in Datei}; 0={nein}
K5	Information Objekte	Sind Informationen über die Rolle/Art der befragten/beforschten Personen/Objekte vorhanden?	FDZ	2={ja, im Abstract}; 1={ja, in Datei}; 0={nein}
K6	Methode	Ist die Forschungsmethode (zur Datenerhebung) genannt?	FDZ, FAIR F.2.	2={ja, in Titel oder Schlagworten}; 1={ja, in Abstract}; 0={nein}
K7	Dateien	Werden die Dateien und deren Inhalt beschrieben?	ergänzt	2={ja, im Abstract}; 1={ja, in Datei}; 0={nein}
K8	Sprache	Ist die Sprache der Daten erkennbar?	ergänzt	2={ja, via Metadaten}; 1={ja, via Titel/ Abstract}; 0={nein}
K9	Offene Formate	Sind die Daten lesbar? (offenes Format)	FAIR R.1	2={ja, komplett}; 1={ja, teils}; 0={nein}

Kriterien ID	Name	Beschreibung	Quelle	Prüfung
K10	Interoperable Formate	Erlauben die Daten eine einfache Bearbeitung?	FAIR R.1.3	2={komplett interoperabel}; 1={teils interoperabel}; 0={nein}
K11	Lizenz	Ist der Datensatz unter einer Lizenz veröffentlicht worden, die eine Nachnutzung erlaubt?	FAIR R.1.1	3={ja, OER}; 0={nein}
K12	Anonymisierung	Wurden die Daten anonymisiert?	FDZ	2={ja, komplett}; 1={ja, teils}; 0={nein}; X={nicht zutreffend}
K13	Transkriptionsregeln	Sind die verwendeten Transkriptionsregeln verfügbar?	FDZ	1={ja}; 0={nein}; X={nicht zutreffend}
K14	Instrumente	Sind Erhebungsinstrumente verfügbar?	FDZ	2={ja, in separater Datei}; 1={ja, innerhalb der Ergebnisse}; 0={nein}; /= {nicht zutreffend}

Grundsätzlich erfolgte die Messung in drei Schritten:

- Prüfung, ob die Information eines Kriteriums in einem dafür definierten strukturierten Metadatenfeld vorhanden ist. Bspw. gibt es ein Feld zu „Datum der Studie“ oder „Sprache“ und enthält dieses Angaben? Definierte Metadatenfelder sind gezielt recherchierbar, sodass angenommen wird, dass diese a) bei der Suche und Filterung großer Ergebnismengen nützlich sind, und b) relevante Kriterien direkt ersichtlich sind und somit die Relevanzbewertung der Daten erleichtern.
- Prüfung, ob die Information eines Kriteriums in einem nicht weiter definierten Metadatenfeld vorhanden ist, wie Titel, Abstract oder Beschreibung. In diesen Metadatenfeldern sind Informationen bspw. über die Methode in der Einzeltrefferansicht eines Datensatzes direkt ersichtlich.
- Prüfung, ob die Information eines Kriteriums in einer der Dateien des Datensatzes vorhanden ist. Somit wäre die Transparenz gegeben, jedoch ist die Information nicht direkt recherchierbar oder in der Trefferansicht ersichtlich. Ein/e Nutzende:r müsste erst im Datensatz und unter Umständen

den in mehreren Dateien nach der Information suchen, was die Prüfung der Nachnutzung erschwert.

Zum dritten Aspekt ist zu sagen, dass gut dokumentierte Datensätze relevante Informationen über ihre Daten in einer Übersicht darstellen, oft gekennzeichnet als *readme*-Datei oder auch Datenmanagementplan (DMP). Diese detaillierten Informationen wurden nur im Hinblick auf die festgelegten Kriterien überprüft, da der Studienfokus auf dem Auffinden von Datensätzen lag. Eine Prüfung der Vollständigkeit und Verständlichkeit von DMP wäre sicherlich ebenso interessant.

4 Ergebnisse und Diskussion

4.1 Sind LIS-Datensätze potenziell gut nachnutzbar?

Tabelle 4 zeigt die Ergebnisse der Untersuchung für die einzelnen Kriterien. Die Ergebnisse pro Datensatz sind online verfügbar (Schobert et al., 2023). Anzumerken ist, dass keine Prüfung der wissenschaftlichen Qualität der Datensätze erfolgte, sondern nur geprüft wurde, ob die Informationen der verwendeten Kriterien ersichtlich sind.

Tab. 4: Ergebnisse der Prüfung der Qualitätskriterien pro Kriterium

Kriterien ID	Name	% des Kriteriums erfüllt
K1	Autor*in IDs	53
K2	Ort Datenerhebung	78
K3	Zeit Datenerhebung	56
K4	Anzahl Objekte	63
K5	Information Objekte	94
K6	Methode	64
K7	Dateien	94
K8	Sprache	8
K9	Offene Formate	84
K10	Interoperable Formate	90
K11	Lizenz	95
K12	Anonymisierung	100
K13	Transkriptionsregeln	80
K14	Instrumente	67

Die Kriterien K5, K7 sowie K10–K12 werden von 90% der Datensätze erfüllt. K5 und K7 beziehen sich auf die Transparenz bezüglich der Zielgruppen bzw. Objekte sowie die Beschreibungen, welche Dateien mit welchen Inhalten der Datensatz enthält. Beides sind relevante Kriterien, um über die Handhabung des Datensatzes als auch auf die Nachnutzung zu entscheiden. Bei der Transparenz schneidet K6 mit 65% und Bezug auf die Nennung der Methode schlechter ab. Ein Grund hierfür ist die Komplexität des Kriteriums (siehe Abschnitt 4.2). Bis auf zwei Datensätze haben alle eine offene Lizenz gemäß dem Prinzip von Open Educational Resources (K11), d. h. Datensätze können verwendet werden, entweder ohne jegliche Einschränkungen (Public Domain) oder unter Nennung der Autorenschaft (CC BY) und/oder Weitergabe unter gleicher Lizenz (CC BY SA). Wenn gemäß Datenschutz notwendig, wurden die Daten zudem anonymisiert (K12).

Auch bei den publizierten Dateiformaten ist festzustellen, dass viele Datensätze in offenen (K9) und auch in der Domäne bekannten und verwendeten Formaten (K10) vorliegen. Einige Datensätze bieten die Daten in verschiedenen gut bearbeitbaren Formaten (wie csv) und gleichzeitig in proprietären Formaten (wie xlsx) an, die dann aber schon formatiert mit entsprechenden Anwendungen geöffnet werden können, oder die gut lesbar sind (wie pdf). Somit besteht für Dritte eine Wahl je nach Priorität.

Lediglich bei den ORCIDs, der eindeutigen Identifikation der Autor*innen, liegt der Prozentanteil mit 53% nicht so hoch. Dies mag in einigen Fällen daran liegen, dass entweder die Datensätze schon länger publiziert sind und eine ID im Repositorium damals nicht angegeben werden konnte. Andererseits wird vermutet, dass einige Datensätze von Studierenden stammen, die nicht langfristig in der Forschung arbeiten und sich somit keine ORCID zulegen.

K3, K4 und K6 sind zu etwa 50% erfüllt. Der Zeitpunkt der Erhebung (K3) war nicht immer genannt, was aber bei manchen Datensätzen zur Beurteilung nicht unbedingt notwendig ist, wie bspw. bei der Erstellung von Vokabularen oder Publikation von eigenen Daten, bei denen das Publikationsdatum als das ausschlaggebende Kriterium angesehen werden kann. Bei Daten aus klassischen Erhebungen wie Umfragen oder Interviewstudien können Erhebungs- und Publikationsdatum sehr wohl weit auseinander liegen, sodass der Zeitpunkt der Erhebung relevant wird. Bei der Angabe der Anzahl der beforschten Personen bzw. Objekte (wie bspw. Onlinedaten) (K4) liegt der Unterschied im Detail: Einige Datensätze haben die Anzahl direkt im Titel oder Abstract, andere in einer Datei bzw. die Anzahl muss aus den Datensät-

zen erschlossen werden. Nur bei drei Datensätzen konnte keine Angabe gefunden werden. Schlusslicht bei den Kriterien ist K8 (Sprachangabe) mit nur 8%. Der Grund hierfür ist die Tatsache, dass es nicht bei allen Repositorien ein Metadatenfeld „Sprache“ gibt – und wenn doch, dies fast nie angegeben wird. Die Sprache mag für viele in der Regel anhand des Titels und Abstracts sofort erkennbar sein. Dennoch gibt es Kontexte, in denen die Sprache der Datendateien nicht eindeutig ist, bspw. bei zweisprachigen Abstracts. Zudem wäre die Sprache in internationalen Repositorien ein wichtiger Filter, um potenziell relevante Datensätze zu suchen, die man als Forscher selbst nutzen könnte. Bei den Unterschieden zwischen den Repositorien lässt sich erkennen, dass die Datensätze an den Repositorien der Hochschulen etwas besser abschneiden als die in den offenen Repositorien Zenodo und Figshare (Tab. 5).

Tab. 5: Erfüllte Kriterien bezogen auf das jeweilige Repository

Mittelwert % über Repo	Anzahl Datensätze	Repo
78	9	Dryad
76	15	HU Berlin
73	4	HAW Hamburg
72	1	Uni Regensburg
67	12	Zenodo
52	3	Figshare

Zusammenfassend lässt sich sagen, dass die meisten Datensätze die festgelegten Kriterien gut erfüllen, der Durchschnitt liegt bei 72%, nur zwei Datensätze liegen unter 50%. Die offenen Lizenzen gemäß den Prinzipien von Open Educational Resources, gut erkennbare Dateibeschreibungen und Inhalte sowie die Verfügbarkeit von verschiedenen gut bearbeitbaren und gängigen Dateiformaten tragen dazu bei, dass die Datensätze verständlich und für potenzielle Nachnutzer gut zugänglich sind.

4.2 Sind die existierenden Qualitätskriterien ausreichend?

Die Prüfung der verwendeten Qualitätskriterien war im Detail herausfordernd, da sich nicht alle Kriterien objektiv eindeutig prüfen lassen. Des Weiteren sind nicht alle Kriterien für alle LIS-Studien zutreffend.

Einen Erhebungsort (K2), der u.a. für qualitative Interviewstudien oder Befragungen relevant ist (z.B. „Studierende welcher Uni wurden befragt?“), gibt es bei der Generierung von Onlinedaten nicht, dafür aber die Quelle wie Twitter als Social-Media-Portal oder das Web of Science. Mitunter ist diese Angabe für die Relevanzbewertung von Datensätzen jedoch zu wenig und eine konkrete Beschreibung, welche Datensätze wie aus den Portalen generiert wurden, wünschenswert. Ein wichtiger Zusatz wäre hierbei auch die konkrete Angabe der Methode bzw. Details zur Datenerhebung (K6). Gerade dieses Kriterium kann mitunter sehr subjektiv bewertet werden, da es in unterschiedlichen Detailierungsgraden beschrieben wird. In der vorliegenden Untersuchung wurden für K6 sehr großzügig Punkte vergeben. Ob die Angaben jedoch in allen Fällen ausreichend für eine Entscheidung zur Nachnutzung des Datensatzes sind, müsste basierend auf dem jeweiligen konkreten Forschungsdesign und der Forschungsfrage geprüft werden. Des Weiteren gibt es Datensätze von Studien, die keine empirischen Methoden verwendet haben wie bspw. Arbeiten zur Erstellung von strukturierten Vokabularen oder Datensätzen, deren Informationen von den durchführenden Institutionen bereitgestellt werden.

Auch die Unterscheidung zwischen der Forschungsmethode (K6) und dem Erhebungsinstrument (K14) gestaltet sich schwierig, bspw. wenn Daten aus dem Internet oder Fachdatenbanken generiert wurden. Hier kann die Methode als Art und Weise definiert werden, wie die Daten zusammengetragen wurden. In einigen Fällen lag Code zum Crawlen der Daten vor, der für K14 als Erhebungsinstrument gewertet wurde. Hingegen existiert bei manuell erhobenen Daten, wie beispielsweise aus Fachdatenbanken, kein spezifisches Erhebungsinstrument. Aus diesem Grund gestaltet es sich schwierig, die Kriterien (K14 und K6) zu prüfen. Eine Klärung und präzisere Definition der Erhebungsinstrumente sind hier nötig. Das FDZ, das diese Kriterien vorgibt, fokussiert sich auf Daten aus qualitativen Studien, die in der Regel ein Erhebungsinstrument wie Fragen des Leitfadeninterviews haben. Ein Kriterium, was unserer Meinung nach fehlt, ist das Analyseinstrument bzw. -daten. Hierunter würden bspw. Codebücher oder Code zur Auswertung von Daten fallen, die von den erhobenen Rohdaten unterschieden werden können.

Auch bei den Beschreibungen der Dateien und Inhalte (K7) können subjektive Maßstäbe angesetzt werden. Dabei ist auch festzuhalten, dass viele Datensätze explizit als Teil einer Publikation gesehen werden. Letztere ist im besten Fall, so wie in vielen der untersuchten Fälle, eindeutig über eine ID verlinkt. Forschende verweisen meist explizit auf die Publikation. Anderer-

seits bedeutet dies auch, dass viele für die Bewertung des Datensatzes relevante Informationen nicht in der Beschreibung des Datensatzes aufgenommen werden, sondern in der Ergebnispublikation zu finden sind. Suchende müssen dann den Umweg über die Publikation gehen. In diesem Fall sehen die datengebenden Forschenden ihre Daten als Teil der Ergebnispublikation und nicht als alleinstehende Datenpublikation. Um den Unterschied eindeutiger zu machen, wäre in den Repositorien eine ersichtliche Unterscheidung zwischen beiden Varianten für den Suchenden wünschenswert. In Bezug darauf mag es dann allerdings Einschränkungen bei der Recherche geben. Denn wenn einige Datensätze qualitativ hochwertiger und womöglich mit Metadatenstandards beschrieben sind, sind diese auch besser auffindbar als andere.

Zusammenfassend lässt sich sagen, dass einige der verwendeten Kriterien gut geeignet sind, um vollständige Angaben bei Datensätzen zu überprüfen. Allerdings erfordern andere Kriterien Anpassungen und klare Definitionen, insbesondere bei Erhebungsinstrumenten und Forschungsmethoden. Es gibt auch Kriterien, die eher auf individueller oder pragmatischer Basis bewertet werden können, was den Vergleich zwischen verschiedenen Datensätzen erschwert. Eine mögliche Lösung besteht darin, methodenspezifische Kriterien zu entwickeln, um diese Herausforderungen zu bewältigen.

4.3 Weitere Aspekte zur Recherche von Datensätzen

Aus den Erfahrungen der Studie lässt sich festhalten, dass es durchaus einige interessante Datensätze aus dem LIS-Bereich gibt. Die durchgeführte Recherche war jedoch nicht erschöpfend und es bleibt festzuhalten, dass die Suche nach geeigneten Datensätzen mit Herausforderungen verbunden ist. Einige dieser Herausforderungen beinhalten fehlende Filter oder eine unklare Indexierung. Es ist nicht klar, ob die Daten einheitlich indexiert sind oder ob Daten durch fehlende Filter möglicherweise nicht gefunden werden. Bieten Repositorien beim Zurverfügungstellen der Daten keine oder nicht genügend Metadatenfelder an, können Forschende dort auch keine Informationen eintragen. Dies erschwert wiederum die Auffindbarkeit von Daten. Auch die Art des Retrievals ist mitunter nicht klar. Zum Beispiel treten auf Zenodo unterschiedliche Ergebnisse bei Groß- und Kleinschreibung der Suchterme auf.

Bei der Suche nach dem Begriff „Informationswissenschaft“ oder dem Ausdruck „Information Science“ ergaben sich zahlreiche Suchergebnisse, die nur bedingt zum Fachbereich Informationswissenschaft passten. Einige ge-

fundene Datensätze wurden daher ausgeschlossen. Dies traf vor allem auf die offenen interdisziplinären Repositorien zu. Aus diesem Grund ergab die Anzahl der gesamten Treffermenge 98, wobei nur 44 Datensätze auch tatsächlich der Informationswissenschaft zuzuordnen waren. Diese Schwierigkeiten in Bezug auf Filter, Indexierung und Retrieval unterstreichen die Notwendigkeit einer weiteren Verbesserung und Standardisierung der Such- und Filterfunktionen in Datenbanken und Repositorien. Nur durch solche Verbesserungen kann die Auffindbarkeit und Zugänglichkeit von geeigneten Datensätzen erleichtert werden. Gerade in dezentralen Strukturen, wie sie gerade über die Nationale Forschungsdateninfrastruktur¹⁰ in Deutschland entstehen, wäre eine einheitliche Beschreibung von Datensätzen aus einem Fachbereich zielführend, um eben diese Daten fachbezogen recherchieren zu können. Bereits angewandt wird hierfür die Nutzung der Dewey-Dezimalklassifikation (DDC) in vielen der untersuchten Universitätsrepositorien (siehe Tab. 2).

5 Zusammenfassung und Ausblick

Die Studie untersuchte, welche LIS-Datensätze von bekannten Hochschulstandorten verfügbar sind und wie sie anhand von Qualitätskriterien zur Auffindbarkeit und Nachnutzung bewertet werden können. Dafür wurden bereits vorhandene Kriterien auf eine Auswahl an Datensätzen angewendet. Die Untersuchung ergab, dass viele Datensätze die Kriterien gut erfüllen, was darauf hindeutet, dass sie verständlich und potenziell gut nachnutzbar sind. Insbesondere die Kriterien K5, K7, K10, K11 und K12 wurden von etwa 90% der Datensätze erfüllt. Darunter sind im Sinne von Open Science relevante Kriterien, die sich auf die Transparenz bezüglich der Zielgruppen bzw. Objekte, die Beschreibungen der enthaltenen Dateien und Inhalte sowie die Verwendung offener Lizenzen gemäß den Prinzipien von Open Educational Resources beziehen. Ebenso relevant ist, dass viele Datensätze in offenen (84%) und einfach zu nutzenden (95%) Formaten vorliegen.

Einige Kriterien wurden nicht gut erfüllt, was einerseits auf die Schwierigkeit zurückzuführen sein kann, die Detailgenauigkeit des Kriteriums zu beurteilen (wie im Fall der Forschungsmethode), andererseits darauf, dass Repositorien teilweise keine entsprechenden Metadatenfelder zur Verfügung

¹⁰ <https://www.nfdi.de/>

stellen (wie im Fall der Sprache) oder bestimmte Kriterien nicht explizit auf Studien anwendbar sind (wie im Fall der Erhebungsmethode). Ein Kriterium, bei dem Verbesserungsbedarf besteht, ist die eindeutige Identifikation der Autor*innen (K1). Nur 53% der Datensätze enthalten ORCID-IDs. Für eine verbesserte Transparenz wäre hier ein höherer Wert wünschenswert, wenngleich auch diskutiert werden kann, ob der Aufwand bei Arbeiten von Studierenden nötig ist. In Bezug auf die Unterschiede zwischen den Repositorien zeigte sich, dass die Datensätze an den Hochschulrepositorien etwas besser abschnitten als die in den offenen Repositorien Zenodo und Figshare. Dies ist auch ein Hinweis auf die Unterschiede beim Forschungsdatenmanagement und den Qualitätskontrollverfahren, die an immer mehr Hochschulen umgesetzt werden.

In der Studie wurden weitere Herausforderungen bei der Recherche nach Datensätzen identifiziert, darunter fehlende Filter, eine Intransparenz hinsichtlich der einheitlichen Indexierung in den Repositorien sowie Unklarheiten beim Retrieval. Diese Herausforderungen unterstreichen die Notwendigkeit weiterer Verbesserungen und Standardisierungen der Such- und Filterfunktionen in Repositorien. Die bislang verteilt vorliegenden Datensätze sollten dringend über ein übergreifendes Suchsystem auffindbar sein. Solch ein auf Standards aufbauendes Findesystem könnte beispielsweise eine Aufgabe des relevanten Fachinformationsdienstes sein. Durch entsprechende Maßnahmen kann dann die Auffindbarkeit und Zugänglichkeit von geeigneten Datensätzen effektiv erleichtert werden. In Kombination mit einer fachspezifischen Anpassung von Kriterien zur Bewertung von Datensätzen und der Schulung der Forschenden zum Teilen von Daten kann erreicht werden, dass qualitativ hochwertige LIS-Datensätze publiziert und nachgenutzt werden.

Danksagung

Herzlichen Dank an Jasmin Khoura-Bodji, die uns bei der Prüfung der Datensätze unterstützt hat.

Literatur

- Ackermann, M. R.; Dietze, S.; Jäschke, R.; Katsanidou, A.; Mathiak, B. (o.J.): Unknown Data – Mining and consolidating research dataset metadata on the Web. <https://gepris.dfg.de/gepris/projekt/460676019?language=en>
- Bambey, D.; Corti, L.; Diepenbroek, M.; Dunkel, W.; Hanekop, H.; Hollstein, B.; Imeri, S.; Knoblauch, H.; Kretzer, S.; Meier zu Verl, C.; Meyer, C.; Meyer-
mann, A.; Porzelt, M.; Rittberger, M.; Strübing, J.; Unger, H. von; Wilke, R. (Hrsg.) (2018): Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung. https://www.ratswd.de/dl/RatSWD_WP_267.pdf
- Bambey, D.; Meyermann, A.; Porzelt, M. (2017): Potentiale der Sekundärforschung mit qualitativen Daten – Ein Workshopbericht. https://www.forschungsdatenbildung.de/get_files.php?action=get_file&file=fdb-informiert_nr-7.pdf
- Bambey, D.; Meyermann, A.; Porzelt, M.; Rittberger, M. (2018): Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung. In: D. Bambey, L. Corti, M. Diepenbroek, W. Dunkel, H. Hanekop, B. Hollstein, S. Imeri, H. Knoblauch, S. Kretzer, C. Meier zu Verl, C. Meyer, A. Meyermann, M. Porzelt, M. Rittberger, J. Strübing, H. von Unger, R. Wilke (Hrsg.): Bereitstellung und Nachnutzung qualitativer Daten in der Bildungsforschung (S. 59–68). https://www.ratswd.de/dl/RatSWD_WP_267.pdf
- Bambey, D.; Rittberger, M. (2013): Das Forschungsdatenzentrum (FDZ) Bildung des DIPF. Qualitative Daten der empirischen Bildungsforschung im Kontext. In: D. Huschka, H. Knoblauch, C. Oellers, H. Solga (Hrsg.): *Forschungsinfrastrukturen für die qualitative Sozialforschung* (S. 63–71). Berlin: SCIVERO Verlag. <https://doi.org/10.25656/01:8468>
- Bishop, L.; Kuula-Luumi, A. (2017): Revisiting Qualitative Data Reuse. In: *SAGE Open*, 7 (1), 215824401668513. <https://doi.org/10.1177/2158244016685136>
- Eckert, S.; Favella, G.; Gärtner, A.; Kirschner, U. (2023): Standardisierter Datenmanagementplan. Philipps-Universität Marburg. <https://doi.org/10.17192/bfdm.2023.2.8458>
- Fecher, B.; Puschmann, C. (2015): Über die Grenzen der Offenheit in der Wissenschaft – Anspruch und Wirklichkeit bei der Bereitstellung und Nachnutzung von Forschungsdaten. In: *Information – Wissenschaft & Praxis*, 66 (2–3). <https://doi.org/10.1515/iwp-2015-0026>
- Feldsien-Sudhaus, I.; Rajski, B. (2016): Digitale Forschungsdaten für die Zukunft sichern: Umfrage zum Umgang mit Forschungsdaten an der TU Hamburg: Auswertung. <https://doi.org/10.15480/882.1326>
- Heck, T.; Rittberger, M.; Schneider, G. (2021): Setting standards – how information centres can help to foster open science. <https://doi.org/10.5281/ZENODO.8362934>

- Heck, T.; Schneider, G. (2021): Quality of open research data in education. Vorab-Onlinepublikation. <https://doi.org/10.5281/zenodo.4672653>
- Kindling, M. (2013): Qualitätssicherung im Umgang mit digitalen Forschungsdaten / Quality assurance of digital research data / La garantie de la qualité des données numériques de recherche. In: *Information – Wissenschaft & Praxis*, 64 (2–3), 137 bis 147. <https://doi.org/10.1515/iwp-2013-0020>
- Lösch, T.; Bayer, S.; Heck, T.; Blask, K.; Perry, A.; Toepfer, R.; Vlaeminck, S. (2022): Forschungsdaten nachnutzen in der Krise: Wie wurden vorhandene Daten während COVID-19 genutzt? [Re-using research data in crisis: How was open data used during COVID-19?]. In: *9. Conference of Society of Empirical Educational Research (GEBF)*.
- Neuroth, H.; Oevel, G. (2021): Aktuelle Entwicklung und Herausforderungen im Forschungsdatenmanagement in Deutschland. In: M. Putnings, H. Neuroth, J. Neumann (Hrsg.): *Praxishandbuch Forschungsdatenmanagement* (S. 537–556). Berlin: De Gruyter Saur. <https://doi.org/10.1515/9783110657807-029>
- Neuroth, H.; Rothfritz, L.; Petras, V.; Kindling, M. (2019): Digitales Datenmanagement als neue Aufgabe für wissenschaftliche Bibliotheken. Vorab-Onlinepublikation. <https://doi.org/10.18452/20680>
- Pampel, H.; Vierkant, P.; Scholze, F.; Bertelmann, R.; Kindling, M.; Klump, J.; Goebelbecker, H.-J.; Gundlach, J.; Schirmbacher, P.; Dierolf, U. (2013): Making research data repositories visible: the re3data.org Registry. In: *PLOS ONE*, 8 (11), e78080. <https://doi.org/10.1371/journal.pone.0078080>
- Piowar, H. A.; Vision, T. J. (2013): Data reuse and the open data citation advantage *PeerJ* 1:e175. <https://doi.org/10.7717/peerj.175>
- Rahal, R.-M.; Hamann, H.; Brohmer, H.; Pethig, F. (2022): Sharing the Recipe: Reproducibility and Replicability in Research Across Disciplines. In: M. Schubotz, S. Behrens, T. Heck, R.-M. Rahal, I. Steinhardt, D. Scholl (Hrsg.): *Bootstrapping the Open Science culture: The fellowship approach*. Pensoft Publishers. <https://doi.org/10.3897/rio.8.e89980>
- Rösch, H. (2021): Forschungsethik und Forschungsdaten. In: M. Putnings, H. Neuroth, J. Neumann (Hrsg.): *Praxishandbuch Forschungsdatenmanagement* (S. 115 bis 140). Berlin: De Gruyter Saur. <https://doi.org/10.1515/9783110657807-006>
- Scherp, G.; Siegfried, D.; Biesenbender, K.; Breuer, C. (2020): Die Bedeutung von Open Science in den Wirtschaftswissenschaften. Ergebnisbericht einer Online-Befragung unter Forschenden der Wirtschaftswissenschaften an deutschen Hochschulen 2019,. Kiel/Hamburg. <http://hdl.handle.net/11108/448>
- Schobert, K.; Heck, T.; Rittberger, M. (2023): Qualität und Nachnutzungspotential von LIS-Forschungsdaten in offenen Repositorien: Datensatz. <https://doi.org/10.5281/zenodo.8360112>

- Tenopir, C.; Rice, N. M.; Allard, S.; Baird, L.; Borycz, J.; Christian, L.; Grant, B.; Olendorf, R.; Sandusky, R. J. (2020): Data sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. In: *PLOS ONE*, 15 (3), e0229003. <https://doi.org/10.1371/journal.pone.0229003>
- Wallis, J. C.; Rolando, E.; Borgman, C. L.; Nunes Amaral, L. A. (2013): If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. In: *PLOS ONE*, 8 (7), e67332. <https://doi.org/10.1371/journal.pone.0067332>
- Watteler, O. (2022): Daten in den Sozialwissenschaften. In: M. Tausendpfund (Hrsg.): *Forschungsstrategien in den Sozialwissenschaften* (S. 225–256). Wiesbaden: Springer VS. https://doi.org/10.1007/978-3-658-36972-9_10
- Wilkinson, M. D.; Dumontier, M.; Aalbersberg, I. J. J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L. B.; Bourne, P. E.; Bouwman, J.; Brookes, A. J.; Clark, T.; Crosas, M.; Dillo, I.; Dumon, O.; Edmunds, S.; Evelo, C. T.; Finkers, R., ... Mons, B. (2016): The FAIR Guiding Principles for scientific data management and stewardship. In: *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>

Citation for this Paper

Error generating LaTeX citation.

BibTeX Citation:

```
@inproceedings{schobert_qualitat_2024,  
  author={Katharina Schobert and Tamara Heck and Marc Rittberger},  
  title={Qualität und Nachnutzungspotenzial von LIS- Forschungsdaten in offenen Repositorien},  
  booktitle={Nachhaltige Information — Information für Nachhaltigkeit},  
  date={2024-04-01},  
  publisher={Verlag Werner Hülsbusch},  
  pubstate={published},  
  tppubtype={inproceedings},  
  year={2024},  
  month={02}  
}
```