

Q: Und dann legen wir direkt los. Also, am Anfang geht's erstmal praktisch um deine Perspektive als Datennutzer auf Sekundärdatennachnutzung sozusagen. Und zwar würde ich da erstmal zum Einstieg gerne wissen, wie oft du in der Vergangenheit Datensätze aus deinem Labor oder anderen Labors wiederverwendet hast und ich würde dich auch bitten, das Ganze so ein bisschen zu quantifizieren, indem du einfach die relative Häufigkeit angibst, mit der du Daten wiederverwendet hast, relative eben zur Erzeugung von Primärdaten. #00:00:40-4#

R: Ja, also ich würde schätzen, ungefähr einmal pro Jahr vielleicht, jede fünfte Studie oder so. #00:00:46-3#

Q: Mhm. Und wofür nutzt du Sekundärdaten hauptsächlich? #00:00:55-8#

R: Ganz unterschiedlich. Also, manchmal sind es eigene Daten, die wir noch mal auswerten, unter anderen Gesichtspunkten. Manchmal sind es Daten von Kooperationspartnern, dass wir da noch eine Zusatzstudie draus machen, weil uns da noch was eingefallen ist, was man eben Spannendes mit den Daten machen kann. Dann schicken die mir die. Und bei uns im *[Forschungsbereich 1]* Bereich gibt's noch den Sonderfall, dass es diese international vergleichenden Schülerleistungsstudien gibt. Die PISA z.B., die von der OECD durchgeführt werden. Da werden die Datensätze online gestellt und da gibt's ein Codebook, das ist 400 Seiten stark. Und das ist natürlich super. Und das ist auch meine Antwort, dass, wenn ihr sowas machen wollt, was ich eigentlich brauche, ist dieses Codebook mit 400 Seiten Details. So einzelne Stichworte helfen mir immer nicht so viel weiter bei solchen Sachen. #00:01:37-2#

Q: Ja, das ist klar, ja. Also, ein ausführliches Codebook wäre dein Wunsch (lacht). #00:01:44-9#

R: Ja, und zwar vor allem, also es gibt ja auch Codebooks, das ist einfach nur eine Liste von Stichworten, das ist wirklich eine Fließtextbeschreibung. Da wird der theoretische Hintergrund, wie die Maße entwickelt wurden beispielsweise, was da, auf welchen theoretischen Hintergrund sich das stützt, die genaue Evaluation der Items, die Testkonstruktion, das wird alles im Fließtext erwähnt, das sind 400 Seiten, das ist so wie 10 Artikel oder so. #00:02:06-1#

Q: Ah ja. #00:02:06-8#

R: Und dann hat man aber auch das Gefühl, dass man die Daten wirklich verstanden hat und die Tests, weil...ja, oft, so, das ist halt das Problem an den Sekundärdaten, es hängt ja oft an den Details, ne? In der Experimentalpsychologie lassen sich ja oft Studien nicht replizieren und dann sagen die Replikatoren: Ich hab's aber genauso gemacht wie die Originalautoren. Und die Originalautoren sagen dann: Nein, es gibt subtile Details, von denen ihr nichts wusstet, die sind, das sind meistens... so viele und das ist so subtil, das lässt sich sprachlich gar nicht beschreiben in den Artikeln und das habt ihr falsch gemacht. Und das ist natürlich immer das Problem, wenn man mit Sekundärdaten von anderen Leuten arbeitet, dass es so eine Art Blindflug ist, da man die Details nicht kennt, ne. Und je mehr Details ich also habe, wirklich im Fließtext zusammenhängend von anderen Menschen als kohärente Geschichte für mich dargestellt, umso eher kann ich den Daten auch vertrauen und kann vielleicht auch versuchen, Subtilitäten der Ergebnismuster zu interpretieren. #00:02:54-6#

Q: Ja, das ist richtig. Okay. Also, zusammenfassend kann man sagen, nutzt du eigentlich schon so, wenn man's klassisch interpretiert, das breite Spektrum auf jeden Fall auch von Sekundärdaten und deren, ja, Nachnutzungsmöglichkeiten, also einmal Re-Analyse und zum anderen aber auch Meta-Analysen stark, ne, also so eine Synthese sozusagen von bestehenden Daten? #00:03:20-0#

R: Mhm (zustimmend). Genau. Vielleicht, was man auch noch, was so ein Seitenaspekt ist, wir machen ja auch viele Meta-Analysen, wo wir ja die Effektstärken von anderen Leuten kodieren und uns teilweise auch Zusatzinformation per Email noch von denen holen. In gewisser Weise ist das ja auch so eine Art Nachnutzung, ne? #00:03:33-1#

Q: Genau. Das ist, genau, das fällt auch darunter. Und diese Informationen, die ihr euch da noch zusätzlich holt von den Forschenden, wäre das auch was, was man schon sozusagen als Metadatum bereitstellen sollte deiner Ansicht nach? Also, damit ihr euch einfach diesen Arbeitsschritt quasi sparen könnt, noch mal in Kontakt zu treten mit den Originalautoren? #00:03:57-0#

R: Ja, klar, wenn die ihre Datensätze vernünftig gelabelt hochladen würden...also, für Meta-Analysen, da will man ja oft nur so die Gesamt-Effect-Size haben oder noch ein paar

Moderatoren dazu, da braucht man's ja...da zählen ja nicht so die subtilen Details irgendwie. Und für Meta-Analysen wäre es auf jeden Fall sehr hilfreich und da würde auch schon so ein größeres System helfen. #00:04:12-8#

Q: Okay. Aber du würdest dir dann, sage ich mal, in diesem sehr umfangreichen und ausführlichen Codebook würdest du dir da dann auch die Effektstärken drin vorstellen sozusagen? Also, dass da wirklich alles drin ist, von Theorie über Hypothesen, Analysen, Ergebnissen. #00:04:33-3#

R: Nee, die...ähm, also, ich will ja...also, die Ergebnisse haben die Leute ja in der Regel publiziert, und ich will ja jetzt eine Re-Analyse mit den Daten machen. Wenn jetzt die Leute ihre Analysen und die Ergebnisse noch mal in ihr Codebook schreiben würden, das wäre aus meiner Sicht relativ sinnlos, weil, ich will ja neue Analysen mit den Daten rechnen. Was ich bräuchte... (...) #00:04:53-6#

Q: Nee, ich meine, wenn du eine Meta-Analyse machst, beispielsweise. Für den Fall. #00:05:01-2#

R: Für eine Meta-Analyse ist es auch immer besser, man hat die Individualdaten und kann direkt damit die Effect Sizes noch mal neu berechnen. Das ist immer besser. Insofern, an den Daten (, an der...) Auswertung von anderen Leuten, die Auswertung würde mich nicht so interessieren, sondern...na, also, wenn du jetzt als...wenn man sich das als Teil vom Artikel vorstellt, eine ausführliche Einleitung, ein ausführlicher Methodenteil sozusagen, und, ja, Ergebnisteil und den Diskussionsteil, den würde ich dann neu und selber machen wollen mit den Daten sozusagen. #00:05:25-7#

Q: Okay. Aber das heißt sowas wie beispielsweise die Bereitstellung von Analyseskript und Aufbereitungsskript sozusagen, wie man von den Rohdaten beispielsweise zu den aggregierten Daten gekommen ist, das wäre schon hilfreich oder würdest du das auch nicht brauchen unbedingt? #00:05:43-3#

R: Doch, das wäre ja Teil des Methodenteils sozusagen, ne. Wie die Daten aufbereitet wurden. Das ist ganz, ganz wichtig, weil durch die Aufbereitung der Daten ergibt sich natürlich auch, wie wurden missing values geschätzt und all solche Details und (...) ganz

wichtig sind (...) auch wieder die Details der Ergebnisse interpretieren zu können für mich.
#00:05:55-2#

Q: Mhm. Okay. Also zusammenfassend quasi dann Codebook plus diese Analyse und Datenaufbereitungsskripts. #00:06:09-6#

R: Mhm (zustimmend). #00:06:06-5#

Q: Ja? Kann man das so...? #00:06:12-7#

R: Mm, also, wenn die im Codebook vernünftig mit Fließtext beschreiben alle Details der Datenaufbereitung, dann brauche ich das Skript im Grunde genommen nicht. Da gibt's ja dann auch fünf, sechs unterschiedliche Programme, die die Leute nutzen können, von denen ich nur zwei oder drei kenne vielleicht oder sowas. Also, eine vernünftige, nachvollziehbare Fließtextbeschreibung hilft mir erstmal zu verstehen, was passiert ist, und das ist das A und O. Alles andere ist dann ergänzend irgendwie. Kann ja auch sein, dass ich die Syntax von den anderen Leuten gar nicht verstehe oder eigentlich ist es ja auch egal, wie genau die was programmiert haben. Weißt du, du kannst in R ja auch dieselbe Sache auf zehn unterschiedliche Arten und Weisen programmieren. Das ist mir gar nicht so wichtig. Sondern mir ist wichtig, wie wurden die missing values geschätzt? – Wenn die das in einem Satz schreiben, dann würde mir das schon reichen. #00:06:50-9#

Q: Mhm. Okay. Gut. Das ist mal eine andere Perspektive. (lacht) Bisher haben die Forschenden meist gesagt, so, ah ja, Analyseskript und Datenaufbereitung...aber so, ja, das ist natürlich für den Forschenden eigentlich einfacher verständlich, ne, wenn er es dann so in der Form von einem kurzen Artikel, sage ich mal, auch präsentiert kriegt eher. #00:07:16-1#

R: Wenn die Leute alle saubere, vernünftige, gut auskommentierte Syntax schreiben würden, wäre es ja auch wieder was anderes, aber tun ja die meisten auch nicht. #00:07:23-3#

Q: Das stimmt. Okay. #00:07:26-3#

R: Na, sonst wär's vielleicht wichtig, dass, wenn ihr so ein standardisiertes System machen wollt, dass man auch Standards da vorgibt, wie so eine Syntax aussehen sollte. Das wäre dann

wieder sehr hilfreich, wenn die Syntax so vorstrukturiert wie ein Artikel oder sowas, dass es da vordefinierte Blöcke gibt und so und halt auch (...) vorgegeben ist, was kommentiert werden muss und was nicht und so, dann würde das sicherlich auch (...), das die Sache noch mehr erleichtern. #00:07:47-2#

Q: Ja. #00:07:45-5#

R: Weil, es ist ja alleine schon so, wenn jetzt, also, wenn ich missing values auf fünf unterschiedliche Arten und Weisen schätzen kann und die Leute entscheiden sich für eine und schreiben das in ihre Syntax rein, dann sehe ich an der Syntax, welche dieser fünf Arten sie genommen haben, aber ich sehe nicht, warum, mit welchem Argument sie sich dafür entschieden haben. Und wenn es kluge Leute waren, dann haben die vielleicht aus ganz bestimmten Gründen auf die vier anderen Methoden verzichtet und diese fünfte Methode genommen. Und da würde ich eben erwarten, dass, wenn jemand eine kohärente Fließtextbeschreibung schreibt wie so einen Methodenteil, dann würde ich erwarten, dass das da drinsteht, aber das würde man in eine Syntax normalerweise nicht reinschreiben, ne. Deswegen fände ich jetzt eben diesen Fließtext hilfreicher. #00:08:18-3#

Q: Ja. Ja. Das geht so ein bisschen in Richtung preregistered report, ne? Das, was du jetzt beschreibst, sag ich mal. (lacht) Also, weil, da gibt's ja dann auch hauptsächlich dann eben diesen Theorieteil und den Methodenteil, ne. #00:08:34-7#

R: Mhm (zustimmend). #00:08:35-8#

Q: Und dass man dann sozusagen sowas ähnliches auch in das Codebook eben integriert und darüber dann sozusagen dem Forschenden erlaubt, die...die Methode und...und auch den Denkansatz sozusagen, der den Daten zugrunde liegt, nachvollziehen zu können. #00:08:53-5#

R: Auf jeden Fall, ja, genau, das geht stark in die Richtung, würde ich auch so sehen. Der Unterschied ist vielleicht ein bisschen, wenn ich einen preregistered report mache, ist das ja eine spezifische Studie, die ich durchführen möchte. Ich möchte ganz spezifische Thesen testen, und die Einleitung und der Methodenteil sind auf die Testung dieser Hypothesen zugeschnitten. Wenn ich jetzt Re-Analysen machen möchte, möchte ich ja genau diese

Hypothesen nicht testen, weil die wurden ja schon gemacht und die wurden ja schon publiziert. Ich möchte ja etwas (...) ganz Neues mit den Daten machen. Und deswegen wäre es eigentlich hilfreicher, wenn jetzt die Einleitung und Methodenteil nicht so geschrieben werden, als würde es jetzt um die Testung von ganz spezifischen Hypothesen gehen wie bei einem preregistered report, sondern, wenn eben die Datenerhebung an sich in ihrer ganzen Breite, mit allen explorativen Variablen und allem, was man sonst noch machen könnte und so, wenn das halt beschrieben wird, ne. #00:09:40-4#

Q: Ah, okay. #00:09:40-4#

R: Also, bei der PISA-Studie ist das halt so. Bei der PISA-Studie haben die irgendwie 180 Variablen bei 180.000 Schülerinnen und Schülern gemessen, und das ist halt so aufwändig, dass man sagt, man kann gar nicht im Vorhinein abschätzen, was für Daten mit diesen Analysen...was für Analysen mit diesen Daten überhaupt möglich sind. Deswegen beschreiben wir die gesamte Methode so ausführlich und umfassend wie möglich, und wir stellen den Datensatz online und dann hoffen wir, dass die Leute so viele originelle Ideen haben wie möglich, damit halt spannende Sachen zu machen. Ne, so wie die das ja auch bei diesem Leibniz-Institut für das...die...wie, wie heißt das? Für die Bildungsverläufe anmerkt. #00:10:07-5#

Q: GESIS. #00:10:10-2#

R: Nee, für die Bildungsverläufe in Bamberg. #00:10:12-5#

Q: Ah. #00:10:12-5#

R: Die haben so einen...auch so einen großen Längsschnitts-Datensatz mit repräsentativen Bevölkerungsdaten, den die online stellen. Und da geht's auch darum, dass so viele Leute wie möglich diesen Datensatz nutzen sollen. Und die beschreiben ihre Methode und dann kann man damit quasi beliebige Sachen machen. Deswegen ist die Beschreibung der Methode noch sehr offen. #00:10:28-7#

Q: Aha. Okay. #00:10:31-6#

R: Allerdings muss man sagen, wir nutzen auch teilweise diese Daten aus Bamberg und die Beschreibung ist grauenhaft. Grauenhaft. Die haben...die haben Codebücher und es sind...es ist fürchterlich. #00:10:43-2#

Q: Und inwiefern fürchterlich? Also einfach unverständlich im Sinne von zu viele Fachbegriffe beispielsweise, die man jetzt so allgemein nicht kennt oder...ja, also warum fürchterlich? (lacht) #00:10:56-1#

R: Ja, es sind eben Listen von irgendwelchen technischen Wörtern. Es sind Stichwortlisten. Aber es ist keine kohärente Darstellung, wann wurde was erhoben, wie beziehen sich die Dinge aufeinander, wie hängen die Details zusammen, aus welchen inhaltlichen Gründen hat man sich für manche Methoden entschieden und gegen andere. Es ist einfach eine ewig lange Liste von Variablennamen und Item-Nummern und wie die Items im Datensatz heißen und so, und irgendwie wie viele Items in welcher Skala sind. Aber warum jetzt diese Items in dieser Skala sind, das steht da nicht. Und das ist ja genau das, was ich wissen wollte. #00:11:22-8#

Q: Ja, ja, ja, ja. Okay, also die beschreiben es quasi so, wie man sich jetzt, sag ich mal, ein klassisches Codebuch vorstellt momentan. #00:11:34-5#

R: Ja, und das ist halt total redundant mit dem Datensatz selber. Ich kann den Datensatz (auch angucken...) und dann sehe ich ja, wie die Items heißen und wie viele...dass da fünf Items sind in einer Skala oder so. Das ist halt genau die Information, die ich nicht brauche, sondern ich bräuchte eben, wie gesagt, so eine zusammenhängende Fließtextbeschreibung, die mir mal die Logik des Ganzen irgendwie aufzeigt. #00:11:51-2#

Q: Ja, ja. Das ist ein guter Hinweis. Wir arbeiten daran. (lacht) #00:11:55-3#

R: (lacht) #00:11:58-5#

Q: Okay. Dann vielleicht noch kurz zum Abschluss von der zweiten Frage: Gibt es noch Methoden der Daten-Nachnutzung, die du kennst, die du aber selber jetzt noch nicht verwendet hast und wo du sagen würdest, okay, da bräuchte ich jetzt vielleicht noch mehr als jetzt beispielsweise dieses ausführliche Codebuch oder eben Analyseskripts, die gut

dokumentiert sind? Was gibt's da, was du vielleicht noch nennen könntest, nennen wolltest?
#00:12:30-7#

R: Puh. Also, ich kann dir sagen, was ich glaube, was nicht funktioniert. Es gibt doch diese Leute in der ABO-Psychologie, die diese Datenbank von Effect Sizes auf vielen Einzelstudien gemacht haben, und dann kann man auf Knopfdruck die Effect Sizes neu gruppieren, um damit immer neue meta-analytische Ergebnisse sich ausgeben zu lassen. #00:12:46-3#

Q: Okay. Hör ich zum ersten Mal, also, kenn ich jetzt nicht. Mm. #00:12:48-7#

R: Ja, auch so Text Mining- oder Data Mining-Sachen, das ist immer sehr schwierig, weil, wie du merkst, bin ich ein Fetischist, was die inhaltlichen Zusammenhänge und die Subtilitäten der inhaltlichen Details angeht. Und wenn man jetzt einfach nur sagt, da wurde Intelligenz gemessen und Alter und ich guck mal, wie das zusammenhängt, aber ich weiß nicht, wie die Altersgruppe sich genau zusammensetzt, welcher Intelligenztest und so, da glaube ich, also, an so automatisierte Systeme, glaub ich, ehrlich gesagt, doch nicht so richtig.
#00:13:11-2#

Q: Okay. Gut. Aber wenn's jetzt darum ginge, überhaupt erstmal Daten zu finden für deine Forschungsfrage, dann wäre doch so ein automatisiertes System, was dir jetzt aus einer Datenbank die...die Sachen raussucht, die jetzt zumindest mal thematisch in Frage kommen könnten, wäre doch dann eigentlich ganz...ganz nett, oder? #00:13:42-3#

R: Ich glaube nicht. #00:13:41-3#

Q: Nee? #00:13:43-0#

R: Ehrlich gesagt nee. Also, wir machen zum Beispiel eine Meta-Analyse darüber, wie [Konstrukt 1] und [Konstrukt 2] zusammenhängen, ja? Und wir haben die Ergebnisse dieser Meta-Analyse, was rauskommen soll, und auch, wie gut die Studie wird, die Qualität der Studie haben wir total in der Hand darüber, wie wir definieren, was [Konstrukt 1] ist und was [Konstrukt 2] ist. Wir arbeiten ja immer mit sehr, sehr schwammigen Konzepten in der Psychologie, ja selbst Arbeitsgedächtnisse, kein Mensch weiß, was in Wirklichkeit das Arbeitsgedächtnis ist, so. Genau das gleiche für [Konstrukt 1] und [Konstrukt 2] ist es

natürlich noch schlimmer. Ja, auch Aufmerksamkeit, es gibt wie viele unterschiedliche Aufmerksamkeiten usw. Ja? Und wenn jetzt in so einer Datenbank irgendeine Effect Size unter dem Label *[Konstrukt 3]* oder *[Konstrukt 2]* oder *[Konstrukt 1]* gespeichert wird, dann besteht eben die Gefahr, je nachdem, welche Theorie ich überprüfen möchte, aus welcher inhaltlichen Perspektive, mit welchen subtilen argumentativen Details, die ich mir dazu überlegt habe, ist es eben sinnvoll, manche *[Konstrukt 2]*tests als *[Konstrukt 2]*tests zu betrachten, andere vielleicht nicht. Ja, was ist mit Tests (zur *[Konstruktsubtyp 1]*...) *[Konstrukt 1]*. Ist das ein *[Konstrukt 2]*tests oder ist das ein *[Konstrukt 1]*tests? Ja, die werden in so einer Datenbank...da hätte irgendein HiWi garantiert sofort gesagt, *[Konstruktsubtyp 1]* *[Konstrukt 1]* ist ein Teil vom *[Konstrukt 1]*-Test, hätte das sofort unter *[Maß des Konstruktes 1]* eingeordnet, dabei ist es definiert als *[Konstrukt 2]* sozusagen. Ja? Und (...wenn eine Datenbank) soooo detailliert wäre, dass es diese Mehrdeutigkeit, dass im Sinne mancher Theorien *[Konstruktsubtyp 1]* *[Konstrukt 1]* vielleicht stärker mit *[Konstrukt 1]* zusammenhängen sollte, im Lichte anderer Theorien sollte *[Konstruktsubtyp 1]* *[Konstrukt 1]* vielleicht eher mit *[Konstrukt 2]* zusammenhängen. Wenn ihr eine Datenbank hättet, die diese semantischen Mehrdeutigkeiten verarbeiten kann, ja, damit könnte ich was anfangen. Aber wenn jetzt irgendein HiWi kommt und nimmt irgendein Maß und sagt, hier ist...jetzt muss ich mich entscheiden, ist das *[Konstrukt 2]* oder *[Konstrukt 1]* oder *[Konstrukt 3]*, ja exekutive Funktion ist irgendwie alles gleichzeitig, so, und der ordnet das dann irgendwo ein. Das ist halt echt Schrott, so. Und damit suggeriert man den Leuten halt auch eine Sicherheit, die gar nicht da ist, weil die Leute nutzen natürlich dann diese Datenbank. Da kommen unerfahrene Doktoranden, klicken auf den Button und sagen, ja, ich hab jetzt alle *[Konstrukt 3]*maße genommen und hab geguckt, wie die mit *[Konstruktsubtyp 1]* *[Konstrukt 1]* zusammenhängen. Die Leute wissen doch gar nicht, was sie da inhaltlich dann im Detail da gemacht haben. Und dann kann man natürlich auch die Ergebnisse wieder nicht interpretieren. Wenn dann was Kontra-intuitives rauskommt, dann muss ich ja wissen, für jede einzelne Studie, warum ist die jetzt drin, warum ist die nicht drin, wie könnte das die Ergebnisse beeinflusst haben. Und das sehe ich (...) bei so automatisierten Systemen oder auch bei so Kategoriensystemen, wenn die vorgegeben werden. #00:16:15-8#

Q: Ja, okay. Gut. Dann ist das natürlich eine große Aufgabe, vor die die uns stellen. (lacht)
#00:16:27-3#

R: (lacht) Viel Spaß! #00:16:27-2#

Q: Ja, danke. Okay, kommen wir dann zur dritten und letzten Frage des ersten Blocks. Welche Art von Daten verwendest du denn hauptsächlich jetzt für deine verschiedenen Nachnutzungsszenarien? Sind es eher so physiologische Daten oder mehr verhaltensbezogene Daten, Videodaten, Textdaten? (...) oder... #00:16:51-2#

R: Das sind quantitative [*Konstrukt 2*]-tests, im Längsschnitt erhoben. #00:16:54-9#

Q: Mhm. Okay. #00:16:54-2#

R: Also halt Daten, im Grunde genommen Lösungsraten, Lösungszeiten. #00:16:58-2#

Q: Mhm. Und nutzt du jetzt auch, sag ich mal, das sind ja die Daten, die du hauptsächlich verwendest oder ausschließlich? #00:17:12-0#

R: Ich hab auch schon mal als Ko-Autor eine fMRI-Studie publiziert, so, aber...könnte ich jetzt nicht selber auswerten. Also, wir haben auch eine Meta-Analyse zu, weiß ich nicht...also, ich kann, nein, ich kann jetzt nicht schwören, dass ich für alle Zeiten nur mit Verhaltensdaten arbeiten werde. So nicht. Aber ich nutze fast nur Verhaltensdaten. #00:17:29-0#

Q: Okay. Und würdest du sagen, dass es jetzt in diesem, also, das Spektrum, sag ich mal, von Verhaltensdaten ist ja relativ breit doch. Und würdest du sagen, dass es da Qualitätsunterschiede gibt zwischen den verschiedenen Kategorien von Verhaltensdaten? #00:17:49-2#

R: Qualitative Unterschiede? Also, von der Datenqualität her oder...? #00:17:52-4#

Q: Von...von der Dokumentation der Daten her. Also, wenn du jetzt, weiß ich nicht, zum Beispiel einmal hast du eine Meta-Analyse zu Intelligenztests und auf der anderen Seite hast du Meta-Analyse zu...sag ich jetzt mal, irgendwelchen Aufmerksamkeitsstudien, wo du dann Reaktionszeiten und Error Rates anguckst. #00:18:14-6#

R: Ja. (überlegt) Wir lesen und publizieren eigentlich hauptsächlich so in APA-Journals oder Journals, die da sehr ähnlich zu sind, die aus...mit dieser Kultur irgendwie zusammenhängen. Das ist schon alles sehr, sehr einheitlich. Bei den Meta-Analysen stellen wir fest, die

wirkliche Grenze ist nicht so sehr zwischen Pädagogik und Experimentalpsychologie oder irgendwie so, sondern die wirkliche Grenze ist zwischen alten und neuen Studien. Wenn man versucht, Studien aus den 70er- oder auch manchmal auch den 80er-Jahren zu kodieren für eine Meta-Analyse, es ist unglaublich, wie viele Details da fehlen und wie schlecht die geschrieben sind. Bei den neuen Studien aus den letzten 10, 20 Jahren ist es in der Regel kein Problem, unabhängig davon, was das Forschungsthema ist. Abgesehen davon, wenn man natürlich in den Bereich der qualitativen Studien reinkommt, da kann man natürlich mit einer Meta-Analyse dann eh nichts mehr mit anfangen. Das ist klar. #00:18:53-2#

Q: Ja, ja, ja, ja. Ja, da wird's schwierig. Gut. Siehst du denn, sag ich mal, bestimmte Ursachen dafür, dass es bestimmte Qualitätsunterschiede gibt in der Dokumentation? Also, wenn...wenn du's zu Beispiel jetzt weißt und vom Hörensagen auch...kannst du das auch gerne äußern, woher diese Qualitätsunterschiede rühren könnten. (lacht) Also, auch wenn's jetzt über verhaltensbezogene Daten hinausgeht, ja. Wenn du jetzt irgendwie gehört hast, okay, physiologische Daten sind beispielsweise total super dokumentiert oder so, und das liegt da und daran, kannst du das an der Stelle auch gerne äußern (lacht). #00:19:42-6#

R: Pfff, keine Ahnung, ich höre immer nur, dass physiologische Daten oberflächlich interpretiert werden von den Leuten, die sie nutzen. Nee, kann ich nichts zu sagen. Aus meiner Sicht ist auch die Qualität der Daten aus den unterschiedlichen Feldern ist eben, wie gesagt, nicht so groß. Weil die Standards, wie so ein Methodenteil auszusehen hat, die sind ja durch den APA Publication Guide auch vorgegeben, das ist doch sehr einheitlich. Ich kann jetzt nicht sagen, wenn man jetzt in der Nachnutzung irgendwie mit fMRI-Daten, EEG-Daten oder Verhaltensdaten arbeiten würde, das kann sein, dass es da große Unterschiede gibt, aber damit kenne ich mich eben nicht aus. #00:20:14-5#

Q: Mhm, okay. Gut. Dann kommen wir jetzt zur Nachnutzung von Forschungsdaten aus Perspektive eines Datengebers. Also, wenn du quasi jetzt Daten bereitstellst, die in einem Repository hochlädst, da würde mich jetzt als erstes interessieren, welche Metadaten du da generell erstmal angibst, wenn du selber einen Datensatz bereitstellst. #00:20:43-6#

R: Ich bin mir nicht sicher, ob ich jemals einen Datensatz hochgeladen habe. Ich glaub, noch nicht. #00:20:46-3#

Q: Ach so. Okay. #00:20:48-3#

R: Ja. Und wenn, dann würde ich mich natürlich an die Vorgaben des jeweiligen Systems halten. Ich denk, das ist ja gerade der Witz, dass das in irgendeiner Weise standardisiert und geordnet ist und dass da jemand sich Gedanken gemacht hat, wie ihr jetzt gerade, was ein vollständiges Beschreibungssystem oder zumindest wie man so nah wie möglich an ein vollständiges Beschreibungssystem rankommen würde, da wären für mich hilfreich klare Vorgaben vor allem auch. #00:21:05-4#

Q: Mhm. Okay. #00:21:09-3#

R: Wobei (hierbei?...) immer die Gefahr besteht, dass die einheitlichen Vorgaben eben dann der Subtilität der einzelnen Studie möglicherweise nicht ganz gerecht werden, ne. #00:21:16-7#

Q: Ja, ich glaub, das Problem wird man nicht lösen. Also, das...bei...es gibt so viele diffizile Unterschiede zwischen den Studien. Also, da ein einheitliches System zu schaffen, was dann auch noch dem Anspruch eines Standards gerecht wird...ich glaub, das, das kann man nicht auflösen (lacht). Also... #00:21:43-7#

R: Ich glaub...ich glaube, das wird dann aufgelöst, wenn die Leute qualitative Fließtextbeschreibungen ihrer Studien hochladen können in so ein System, und irgendwann wird die künstliche Intelligenz der Sprachverarbeitung so weit sein, dass die künstliche Intelligenz diese Artikel lesen und verstehen kann. Und dann kann mir die KI sagen, okay, da hat jemand eine Aufmerksamkeitsstudie genannt, hat aber (...) hat aber das Wort "Aufmerksamkeit" nicht verwendet, weil der irgendwelche Spezialmaße verwendet hat, aber ich verstehe hier mittels Textanalyse, dass das eigentlich das ist. Und dann schlägt die Studie mir das vor. Also, das wäre eigentlich das ideale System. Aber so weit ist man... #00:22:13-6#

Q: (...) Ja, genau (lacht). Nee, aber mit...aber ich würde sagen, also selbst mit den Mitteln, sag ich mal, der semantischen Suche und Ähnlichem kann man das heutzutage, glaub ich, noch nicht so lösen. Und das Problem ist an der Stelle, denke ich, auch, dass man die Forschenden nicht unbedingt davon überzeugen kann, dass eine Bereitstellung der Daten in dieser Form...ja, also dieser wirklich sehr ausführlichen Form...dass das, ja, ihnen einen...einen

Mehrwert bringt. Also, da müsste man, glaub ich, am Anfang erstmal enorme Überzeugungsarbeit leisten, dass diese Form der Dokumentation wirklich zu einem...zu einer effizienteren und effektiveren Verarbeitung von Sekundärdaten führen kann, bevor das gemacht wird im großen Stil. #00:23:08-5#

R: Das ist sicher so, ja. Aber in der Regel ist es ja so, dass ich die Daten ja eh für irgendeine Studie erhebe. Und diese Studie publiziere ich und die hat eh schon eine Einleitung und einen Methodenteil. Und die Frage ist, kann ich die nicht einfach nehmen? Oder was müsste man da ändern oder noch hinzufügen oder müsste man das standardisieren oder so? Wie müsste man das übersetzen in so ein System? Und dann ist es vielleicht nicht mehr so viel Arbeit. Weiß ich nicht. #00:23:27-8#

Q: Mhm (zustimmend). Ja, also wir haben tatsächlich einen ähnlichen Ansatz, den wir (lacht) im Projekt fahren oder fahren wollen, sag ich mal. Weil, wenn wir jetzt mal zur letzten Frage kommen, da geht's mir darum, dass du sozusagen mir mal mitteilst, wie so ein Metadatenstandard in einer idealen Welt aussehen würde und was so die wichtigsten Informationen darin wären. Und das Denkschema für diese Frage ist der bereits von dir schon genannte Journal Article Reporting Standard von der APA, ne. #00:24:03-1#

R: Mhm (zustimmend). Mhm. #00:24:04-2#

Q: Und, ja, also letztendlich hast du's ja schon größtenteils beschrieben, also, was da so deine Vorstellungen wären, ne. Aber das ist eigentlich so genau dieses Denkmodell, ne. Also, dass man halt sagt, okay, man nimmt einen Standard, der schon in den Köpfen existiert und den die Leute eh nutzen, ne, wenn sie (...) ihren Artikel schreiben, und baut darauf sozusagen den Datendokumentationsstandard auf. #00:24:28-2#

R: Ja. Aber, wie gesagt, das Problem, das ich sehe, wenn man jetzt sich anschaut, was gibt die APA vor, wie so ein Methodenteil geschrieben werden soll beispielsweise... Was bei dieser Replikationskrise bei jeder einzelnen...bei jedem einzelnen Experiment, das sich nicht replizieren lässt, sagen die Autoren wieder: Es gibt Details der Durchführung, die sind so subtil, dass wir sie im Methodenteil nicht beschreiben konnten. Jedes einzelne Mal, ich hab das hunderte Male in den letzten zwei Jahren gehört. Und das ist (natürlich schlecht...), weil, wenn jetzt Leute eine Studie versuchen zu verstehen und sie verstehen sie nicht, ist es ja das

eine Problem. Aber wenn ich jetzt Daten von anderen Leuten nehme und ich verstehe sie nicht und ich mache aber eine neue Publikation draus, die dann womöglich noch in einem peer reviewed Journal erscheint, dann ist das natürlich ein großes Problem, weil sich die Fehler dadurch dann multiplizieren sozusagen. #00:25:11-6#

Q: Ja. #00:25:09-1#

R: Und deswegen würde ich sagen, wenn...also die, die Original-Autoren, die sich ein Paradigma ausgedacht haben und die die Studie zuerst ausgeführt haben, die wissen ja am besten, worauf es ankommt. Und die sollen halt diese Subtilitäten dann im Fließtext so gut beschreiben, wie es halt geht. Und wenn es dann halt Sachen gibt, die sind so subtil, dass man die gar nicht mehr per Text ausdrücken, gut, das ist halt dann halt Pech, so. Was soll man dann machen? Das lässt sich dann nicht ändern. Aber, genau, aber deswegen habe ich halt große Bedenken, wenn es so um vereinfachte Systeme geht, so Metadaten (im Umfang...) von zehn Kategorien und in jeder Kategorie kann...hab ich drei Auswahlmöglichkeiten oder sowas. Ich glaube, das führt total in die falsche Richtung. Ich halte das für ganz...nicht für ganz gefährlich, aber ich denke nicht, dass das nützlich ist. #00:25:52-6#

Q: Okay. Gut. Das ist eine sehr gute Perspektive, muss ich sagen. Also, weil das sich, sag ich mal, zumindest, ja, zu großen Teilen auch mit meinen Ansichten deckt, weitestgehend. #00:26:13-3#

R: Mhm (zustimmend). #00:26:15-2#

Q: Weil ich nämlich auch den Eindruck habe, dass so dieses Automatisierte doch recht schwer zu machen ist einfach an der Stelle. #00:26:20-2#

R: Mm (zustimmend). #00:26:22-8#

Q: Aufgrund der vielen diffizilen Unterschiede, die es da eben gibt. Und dass man eben wirklich auch eben versuchen muss, den...den gesamten Forschungsprozess sozusagen abzubilden in dem...in der Datenbeschreibung. #00:26:36-2#

R: Mhm (zustimmend). #00:26:35-9#

Q: Ja, und da müssen wir jetzt halt hinkommen, dass wir das irgendwie hinkriegen., dass wir da ein System haben, wo die Leute auch mit zurechtkommen und was sie auch möglichst einfach dann in ihren Arbeitsalltag integrieren können. Ich denke mal, das ist das A und O, damit es auch umgesetzt wird letztendlich. #00:26:53-2#

R: Mm (zustimmend). Also, was sicherlich sehr hilft, (...) so aus pragmatischer Sicht für mich ist, wenn ich jetzt in so einem...so einer Online-Datenbank Daten von anderen Leuten finden würde, die vielleicht nützlich sind für mich, wenn ich direkt auf irgendwelche Volltext-Links klicken könnte, um mir die Methodenteile der bereits publizierten Studien mit diesen Daten durchzulesen. Weißt du, das würde ja schon in die Richtung gehen, dass ich eben an eine Fließtextbeschreibung rankomme (...), was die Leute gemacht haben. Das ist das eine. (überlegt) Ich hatte noch eine Idee, was war denn das? (denkt nach) Ah, ich komm grad nicht drauf. (denkt weiter nach) Ja, sorry, ist weg die Idee. #00:27:34-1#

Q: Okay. Aber das, ja, das ist auch eine Idee, die kam jetzt auch schon mal aus [*Land 1*]. (lacht) #00:27:38-6#

R: Mhm. #00:27:39-2#

Q: Da wird das auch schon zum Teil so praktiziert. Und die arbeiten da auch in einem Projekt, dass Daten und...und Artikel sozusagen verlinkt werden, ne. Dass die beide sozusagen eine...eine DOI haben, die...die strukturell miteinander verbunden ist. Also, dass sich das einmal DOI-Link und dann Unterstrich-data und dann DOI-Link-Unterstrich-article... Und dass man darüber sozusagen die verschiedenen Stücke miteinander verlinkt und Gleiches dann auch zum Beispiel für...für...für Versuchsmaterialien macht und Ähnliches, ne. Das kann ja auch hilfreich sein. Wenn die Leute jetzt irgendwie einen neuen Fragebogen entwickelt haben oder Ähnliches und wenn sie den dann halt noch mal separat hochladen oder sie haben einen bestehenden Fragebogen genutzt und verlinken dann dazu. Dass man einfach immer zu den verschiedenen Informationen hingeleitet werden kann dann. #00:28:34-4#

R: Ja. Mir ist wieder eingefallen, was ich noch vorschlagen wollte. Nämlich kam das, glaube ich, noch nie vor, dass wir Daten nachgenutzt haben von Leuten und das geschafft haben, ohne die Leute zu kontaktieren. #00:28:45-0#

Q: Mhm. #00:28:43-1#

R: Weil, es waren dann immer für unsere Studie, aus unserer neuen Sicht von dem, was wir neu machen wollten mit den Daten, waren dann doch immer Sachen relevant, die eben die ursprünglichen Autoren nicht auf dem Schirm hatten. Sonst wäre es ja auch nutzlos, jetzt noch mal eine Re-Analyse mit den Daten zu machen. Und wir mussten dann immer noch mal nachfragen nach Details, die sie in ihrem Artikel nicht beschrieben hatten, weil es ihnen eben um andere Aspekte des Datensatzes ging. Das heißt, wenn es da irgendeine Art von, ich will nicht sagen Social Media Integration gäbe, ja, aber irgendwie, also am (...). Also ich nutze halt viel, ich kenne viele Leute, die ResearchGate sehr intensiv nutzen mittlerweile, wo ja auch (Paper...) immer hochgeladen werden und wo Datensätze auch hochgeladen werden. Und da ist es ideal, weil da brauche ich nur auf einen Button klicken und kann die Leute direkt danach fragen. Hey, du hast da die Daten hochgeladen, ich würde die so gerne zitieren, (was ist da...?) los noch mit der Datenerhebung gewesen und so. Und ich meine, was halt auch immer sehr hilft, merke ich auch bei der Datennachnutzung, ist halt, wenn man die Leute persönlich kennt, ne. Ich habe Kooperationspartner (...), die kenne ich persönlich, und die geben mir ihre Daten. Die geben mir auch alle Details. Die schicken mir auch alle Fragebögen. Die sagen mir auch ehrlich, was schiefgelaufen ist bei ihren Datenerhebungen. Und das ist...sind genau die Details, die ich wissen müsste eigentlich. Also, der beste (...) wäre halt kleine Konferenzen, um die Daten zu organisieren, wo die Leute sich treffen.

#00:29:53-3#

Q: Ja. Dass es so ein...so ein...ja, eher persönlicher Rahmen ist dann. #00:30:01-4#

R: Ja, oder Social Media, dass man zumindest die Leute leicht kontaktieren kann oder so.

#00:30:05-4#

Q: Ja. Ja, wobei ich da auch festgestellt habe, das ist manchmal echt schwierig, ne. Also, ich hatte jetzt letztes Jahr so ein systematic review gemacht auch und wo ich dann auch verschiedene Leute über ResearchGate angeschrieben habe, ob sie mir ihren Artikel schicken könnten und ist ja total interessant und, ne. Und da hat es streckenweise dann echt Monate gedauert, bis die dann mal den Artikel geschickt haben. Manchmal haben sie ihn auch gar nicht geschickt. (lacht) Und, ja, weiß ich nicht, da ist wahrscheinlich dann wirklich so diese...diese Möglichkeit zur Face-to-Face-Interaktion noch zielführender. #00:30:42-0#

R: Mm. #00:30:43-3#

Q: So auch, wenn man so gerade an so...so kleine Niedlichkeiten ran möchte. (lacht)
#00:30:44-7#

R: Mhm, ja. #00:30:49-8#

Q: Könnte ich mir vorstellen so. #00:30:51-8#

R: Mhm (zustimmend). #00:30:51-4#

Q: Gut. Dann vielen Dank für deine Zeit und für diese tollen Ideen und Einblicke auch in deine Arbeit. #00:31:03-0#

R: Bitteschön. Gern geschehen. Wünsche ich euch viel Erfolg für euer Projekt! Habt ihr euch ja viel vorgenommen. #00:31:08-1#

Q: Ja, ja, es ist...für drei Jahre ist es stramm. (lacht) Ja, wir hoffen einfach auf die weitere Kooperation von Forschenden, dass wir da einen Standard nachher am Ende haben, der allen Ansprüchen gerecht wird und dass die Leute dann auch wirklich damit was mit anfangen können und das eben auch was bringt sozusagen in der Steigerung der Effektivität ihrer Arbeit. Das ist ja so ein bisschen das Ziel, ne, dass das wirklich Mehrwert hat letztendlich für den Forschungsprozess, für den Arbeitsprozess. #00:31:40-2#

R: Mm (zustimmend). #00:31:41-1#

Q: Jo. Schauen wir mal, ne. #00:31:42-3#

R: Ja, genau. #00:31:42-2#

Q: Ob wir's hinkriegen. Gut. Dann wünsche ich dir noch einen schönen Tag. #00:31:50-0#

R: Dir auch. Tschüss! #00:31:47-2#

Q: Danke, tschüss! #00:31:51-8#