

Cite as: Kalová, Tereza. (2020). Metadata for Research Data: A Needs Assessment in The Sciences Interview Transcript Dataset [Data set]. Zenodo.

<http://doi.org/10.5281/zenodo.3897321>



This dataset is licensed under the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) License.

I: Interviewer

B: Befragter/Interviewee

1. **Wissenschaftler G – Teiltranskript, persönliches Interview (10.05.2019)**
2. **I:** [...] Zum Anfang, kannst du dich bitte vorstellen und dein Forschungsgebiet kurz beschreiben.
3. **B:** Gerne, *Wissenschaftler G* mein Name. Ich arbeit' zurzeit in, ähm, bei [einer Universität in Wien]. Hab davor, also das jetzt seit circa drei Monaten. Und davor hab' ich acht Jahre in Stockholm gearbeitet [...]. Dort hab' ich angefangen zu forschen, das war gleich 2012 oder 2013. Ich glaub' 2014 war ich dann PhD-Student dort offiziell mit einem Forschungs-, also Auftrag und Plan und hab den abgeschlossen August 2018, da war ich dann fertig. ja.
4. **B:** Und was hab' ich da beforcht? Ähm, man kann das grob zusammenfassen als die Effekte von Zigaretten, elektronischen Zigaretten und dem schwedischen Snüs. Ähm, Effekte, also Gefäßeffekte und so (unv.). Ähm, meine Studien waren drei Expositionstudien, wo ma' gesunde Probanden, ähm, untersucht haben, also, die halt randomisiert werden zu, sie rauchen eine Zigarette oder sie rauchen keine Zigarette oder sie rauchen eine elektronische Zigarette oder keine elektronische Zigarette. Beziehungsweise Snüs, genau dasselbe, sie nehmen Snüs, sie nehmen kein Snüs und welche Effekte hat es in diesen Probanden auf die Gefäße. Da haben wir verschiedene Methoden verwendet, wie wir das messen. Wir haben Blutproben abgenommen und geschaut, also, ist da ein schädigender Effekt auf die Gefäße. [...]

5. **B:** Die Daten sind noch nicht publiziert, aber die sind in meiner wissenschaftlichen, ähm, Arbeit, so ein Kompendium hab' ich zum Schluss schreiben müssen, dass ist dann drinnen und dem ist es publiziert und das ist öffentlich auch. Ähm, ich kann sehr lang reden über die verschiedenen Methoden, ja, die ma' verwendet haben. Das waren sehr spannende Ergebnisse. [...] Das hab' ich dann in Stockholm gemacht und wie ich da fertig war mit dem PhD, war das dann für mich abgeschlossen. [...] Und dann [an dieser Institution] werd' ich mir neues Forschungsgebiet dann suchen.
6. **I:** OK, alles klar. Du hast es jetzt schon kurz angeschnitten aber kannst du jetzt nochmal kurz definieren, was in deiner Disziplin als Forschungsdaten gilt?
7. **B:** Uuh, was sind Forschungsdaten. Also Forschungsdaten, äh, das ist eigentlich a schwierige Frage, weil jetzt [...] in Schweden, ich habe ja a bisschen Kontakt gehabt mit Forschungsdaten, die von, ähm, qualitätsichernden äh Datenbanken rauskommen. Und ich muss ehrlich sagen, ich weiß nicht genau wie das in Österreich gehandhabt wird. Also, in Schweden ist das so, Patienten werden oft in riesengroße Register eingeschlossen. In Schweden ist das äh äußerst populär und sehr digitalisiert. Ähm, meiner Meinung nach deutlich mehr als in Österreich. Ähm, da werden Patienten mit den, mit großen und wichtigen Erkrankungen automatisch in sogenannte Qualitätsregister eingeschlossen, relativ automatisch. Zum Beispiel für äh Herzinfarkte, was eine riesen Patientenpopulation ist einfach, ja, da haben alle Krankenhäuser in Schweden sind ja eingeschlossen in dieses Qualitätsregister. Das heißt, Daten fließen da automatisch vom Patienten ein dieses Register. Ähm, und da wird alles Mögliche also aufgenommen, zum Beispiel wann ist der Herzinfarkt gewesen, wie groß war er, äh welche Risikofaktoren gibt's bei diesen Patienten, Ergebnisse von Blutproben etc. etc.
8. **B:** Das ist an sich nur eine Qualitätsdatenbank. Jetzt weiß ich nicht inwieweit, ich glaub' Patienten müssen nicht mal was unterschreiben dafür, also dass sie da aufgenommen werden, das ist für die interne ähm, also nationale vielleicht auch für die nationale Qualitätssicherung, ähm, wie gut die Krankenhäuser sind. Also, meistens wird, es wird dafür verwendet, dass Krankenhäuser untereinander verglichen werden. Ähm, Patienten und miteinander verglichen aber du kannst nämlich aus diesen Daten Studien machen, ähm und das wird dann, meiner Meinung nach, eine (unv.) Studiendatenbank. [...] Forschungsdaten. Daraus werden dann Forschungsdaten extrahiert aber das ist dann eigentlich, es sind dieselben Daten, du

musst dann halt einen Antrag stellen bei der Ethikkonvention mit einer Fragestellung zu deiner Forschung. Und dann darfst du Daten ähm aus diesem sehr riesen Datenbanken einfach entnehmen und zur Forschung verwenden und dann sind's plötzlich Forschungsdaten. Forschungsdaten, ja. Das ist a mal das Eine was mich jetzt interessiert hat. Also jetzt weiß ich nicht, das weiß ich nicht, wie es in Österreich funktioniert, ja. Ähm, für mich waren Forschungsdaten dann vorhanden, ja. Ich hab' eine Frage, eine konkrete Fragestellung gehabt, zu einer Studie, ja? Ethikkommission durchgegangen und hab' dann eben vorgeschlagen also mit einer Power-Berechnung, also wie viele Patienten ich brauche für diese Fragestellung und welchen statistischen Methoden, dass ich halt diese Frage beantworten kann, könnte ... Denn Patienten sind inkludiert, es werden. Und dann stehen schon die ersten Forschungsdaten, würd' ich sagen, also bei der Aufnahme in die Studie zum Beispiel Baseline äh Werte also, wie alt ist der Patient, ähm, männlich weiblich etc. Also Namen waren in dem Fall nicht interessant. [...]

9. **B:** Und dann je nach äh Methode werden dann diverse Daten produziert. Das können ähm Biodaten sein von Blutuntersuchungen. Es können ähm Befunde oder Werte sein Spirometriemessungen oder anderen Methoden, ja. Das generiert halt extrem viel Daten. Und das wird dann auch den äh Forschungsprobanden dazu geordnet. Und ja, wenn die Studie dann zu Ende ist. Also, dann hammer den zum Beispiel der Exposition A und B. Wir haben eben in beiden Expositionen Forschungsdaten eben generiert, je nachdem welche, was die Fragestellung war. Und dann ist die Forschungsdatenbank für mich fertig. Das sind für dann mich die Forschungsdaten, mit den ich arbeite.
10. **I:** Sag dir der Begriff Metadaten irgendetwas?
11. **B:** Metadaten? ... Ähm, jetzt nicht ...
12. **I:** Was würdest du dir darunter so vorstellen? Auch wenn du vielleicht nicht genau weißt.
13. **B:** Metadaten ... ähm, ich kenne ähm also Metaanalysen, ja? ... Also eine Zusammenfassung aus verschiedenen Studien, also die halt zusammengefasst werden und in diesem Thread Plotter da reinkaut werden. Und das ist dann für mich eine Metaanalyse. ... Kann noch falsch liegen.

14. **I:** Es gibt keine falsche Antwort.

15. **B:** Aha, sehr schön.

16. **I:** Ich hab' hier eine, so eine kurze Definition. Es gibt ganz viele Definitionen von dem Begriff Metadaten. Kannst du sie bitte kurz lesen.

17. **B:** [...] OK, das ist interessant. Das ist einfach eine Beschreibung der Daten also eine externe Datenbank über die Daten, die wir zum Beispiel in vier Studien, fünf Studien produziert haben. So versteh' ich das vielleicht so richtig, so. So würd' ich das verstehen.

18. **I:** Ja, wenn du denn selber wissenschaftlich tätig bist oder warst, hast du sowas, solche Beschreibungen, solche Metadaten für die Daten erstellt?

19. **B:** Nein.

20. **I:** Nein. Und hast du da irgendwie Erfahrung damit, dass du z.B. die Daten aus dieser Datenbank dir geholt hast. Gab es da sowas in dieser Richtung?

21. **B:** Nein, also ich selber ähm. Oder, oh, würde das auch gelten, also wenn wir zum Beispiel, wir müssen ja kata/ also katalogisieren Blutproben, eingefrorene Blutproben. Hmm ähm, da brauchen wir natürlich, das/ sie sind natürlich nicht Daten aber sie sind trotzdem Biodaten oder Blutproben für mich, ja. Und die müssen deutlich ja beschrieben werden, also, wo sie sind und ähm wie sie zuzuordnen sind, ja. Das haben wir natürlich geführt und das wird noch immer geführt. Also das ist vielleicht, ich weiß jetzt nicht, ob das jetzt Metadaten sind, aber das ist für mich eine Kategorisierung oder Katalogisierung dieser Daten, die wir produziert haben, die immer noch vorhanden sind. Und das ist aber in einer wunderbaren Form geführt worden, in ein/ auf Papier (lacht). In einem Büchlein liegen die im [...] Krankenhaus, ja. Und falls man ... Zugriff auf diese Biodaten bräuchte, ja? Müsste man hingehen und in dieses Büchlein schauen, wo in welchem Kühlschrank oder welcher Gefriertruhe eigentlich, das ist minus 80 Grad Gefriertruhen, wo die sind, wo die eingeordnet sind, ja. Und beziehungsweise das ist immer noch anonymisiert, das heißt, du weißt nur, sie gehören zu dieser Studie, sie gehören zu dieser Patienten-ID und in dem Fall zu welcher Exposition sie gehören und zu welchem Zeitpunkt sie

entnommen worden sind. Also diese Daten hammer, aber das ist nicht digital geführt.

22. **I:** OK. Und diese Daten in diesen Datenbanken, diese Patientendaten irgendwie, wie du aus Schweden kennst. Gab's da irgendsowas in diese Richtung? So eine Beschreibung von diesen Daten, sodass du wusstest, was ist das eigentlich für Daten, dass man da hat.

23. **B:** Also [die Universität], das ist dann ... gleich 2017 oder 18 haben eben gestartet eine riesen, generell, allgemeine Biodatenbank oder Biobankhotel oder wie man das nennt, ähm. Wo, also ein Service für Forscher [an der Universität], die eben, wenn sie so eine Datenbank anlegen, dass sie dann schon im Vorhinein mit dieser Einrichtung Kontakt aufnehmen, dass das korrekt katalogisiert wird, oder Metadaten eben erstellt werden. Ähm, das wenn dann ähm Forschungsdaten erhoben werden, die werden, (unv.) von Blutproben, das können auch Gewebeproben sein, die eben in dieses ähm Hotel gebracht werden, katalogisiert werden und das wird dann digital äh auch gespart, ja.

24. **B:** Und ... dann wird das eben zentral aufbewahrt und man kann dann also als Forschergruppe externe Anfragen stellen, also wo befinden sich diese Daten. Also man weiß, also, dort sind sie irgendwo, aber man stellt irgendwann dieser Gruppe eine Frage, "wir bräuchten wieder Daten von so-und-so einer Studie und die äh holen es da einfach raus aus diesem riesen Hotel, ja. Also das besteht schon, ähm, wir haben es nicht gemacht, weil das einfach ur teuer war. Noch bei uns. Und äh.

25. **B:** Wir haben kleine Studien gemacht mit 15 Probanden und das sind nicht so viele Analysen angefallen. Ähm, und es war auch nicht so etabliert dann noch. Ich glaub' das ist auf jeden Fall im Werden. So jetzt große Studien, die äh viele Patienten einschließen, also 500.000 oder mehrere Patientenkohorten-Studien, ähm, die greifen sicher auf diese Daten, also dieses Service da zurück.

26. **I:** Hast du denn in Schweden in einem Team gearbeitet an diesen Forschungsprojekten?

27. **B:** Ja.

28. **I:** OK, haben da diese Metadaten irgendwie, diese Beschreibungen von

Forschungsdaten, irgendeine Rolle gespielt? Bei auch, in der Arbeit im Team.

29. **B:** Sag'n ma so, anfangs natürlich leider ... nein. Man hat sich, glaub' ich, nicht so viele Gedanken drüber gemacht außer dass man bei der Ethik, wenn man Ansucht, man muss den sagen, da wird eine Bioatenbank erstellt, ja. Aber in dem Sinn verlangen die nicht, dass Metadaten dazu erstellt werden. Sie verlangen aber jedoch schon, dass wer dafür verantwortlich ist. Ähm, aber es wird jetzt nicht spezi/, es ist nicht wichtig, also. Die Ethik muss ja korrekt sein und es müssen alle Bedingungen erfüllt werden. Ähm, und das ist halt wichtig. Also, dass eine (unv.) Datenbank erstellt wird und wer dafür verantwortlich ist. Aber es gab niemals ein ... in der Auflage, dass Metadaten erstellt werden. (unv.) Deswegen hat sich die Frage nicht wirklich gestellt. Ich habe einfach das übernommen, wie mein Kollege das gemacht hat, eben mit diesem Büchlein.
30. **I:** Alles klar. Das heißt auch zum Beispiel für diesen Ethikantrag. Ihr musstet da keine, irgendwas ausfüllen, wo ihr sagt, was für Daten da erhoben werden?
31. **B:** Doch. [...] Die Daten, also was wir produzieren, muss natürlich vorgestellt werden. Ähm da müssen auch diverse ähm. Wie sagt man das schön auf Deutsch. Also, es müssen Richtlinien eingehalten werden, zu den erhobenen Daten. [...] Wir müssen natürlich angeben, was wir ähm produzieren oder was wir entnehmen. Und wo das aufbewahrt wird, wie ich gesagt hab', und wer eigentlich dafür verantwortlich ist. Dann kann man sich indirekt denken, der dafür verantwortlich ist, sollte eigentlich wissen, wo das ist und damit er das weiß, sollten vielleicht Metadaten entstehen.
32. **I:** Hast du irgendwann mal, beziehungsweise du hast es schon ein bisschen angeschnitten, Forschungsdaten von anderen verwendet? Die andere erhoben haben sozusagen.
33. **B:** (unv.) Ja, das war aber auch innerhalb, also, wir waren zwei Gruppen. Eine Gruppe in Stockholm und eine in [einem anderen Krankenhaus]. Und ähm ich hab' dort in diesem Krankenhaus [...] eine Studie gemacht, ja. Und wir haben dann zwei Studien zusammengelegt, einfach weil es dasselbe Setting war. Das heißt, ich habe aus einer Vorstudie, die dort schon gemacht worden ist, Daten übernommen, ja. Ähm, das heißt, es war eine andere Person, die das gemacht hat, ja. Aber ich habe die Daten da übernommen, ja. Ähm, und das gemeinsam ausgewertet.

34. **I:** Was war da deine Erfahrung mit den Metadaten, mit irgendeiner Beschreibung von dieser Daten, die du da übernommen hast.
35. **B:** Das war ... furchtbar mühsam. Ähm ... boah, ja, wie hat das angefangen? Also, es hat, ich hätt' zum Beispiel Patientendaten gebraucht und die Frage war, wo sind die a mal? Ja, ähm. Also man musste sich da durchfragen bei diversen Personen, Forschungsschwestern zum Beispiel oder dortige lokale Professoren, wo könnte das sein. Das war vielleicht auf einer Data-Stick irgendwo, oder es war bei irgendeinem auf einem Computer.
36. **B:** Ähm, im Endeffekt habe ich's dann gleich gefunden auf einem Server, ja, wo man auch nicht, also im Krankenhaus, wo die Daten dort gesch/ waren. Gott sei Dank. Und die Daten waren dann so abgespeichert, also ich habe ein paar Daten gebraucht, ja, aber man speichert zum Beispiel ganze Protokolle ab. Man macht eine Untersuchung und bekommt da extrem viel Output. Und von diesem Output interessieren mich zum Beispiel nur vier Variablen oder fünf. Ähm, das hat man so gemacht, also man untersucht dann Patienten, ja, und schreibt dann live, also man bekommt den Output, ja man schreibt diese fünf Sachen rauf, das wird zum Beispiel in eine Excel-Tabelle übertragen, ja. Aber man muss immer die Rohdaten sparen. Das heißt dieser riesen Output wird dann gespart, ja. Jetzt hab' ich dann gesagt, na ja, diese Zettel hab' ich dann bekommen. Ich hab' diese, ja genau Zettel waren das nämlich auch. Es war eine Mischung aus Zettel und Excel-Files. Und wo ich, der Fehler war, z. B. was auch passiert, Übertragungsfehler, eh, dann muss ich auf die Zettel zurückgehen. Und wenn die Zettel nicht lesbar sind, muss man zurückgehen auf die Rohdaten. Das muss alles immer gespeichert werden und die Frage ist immer wo. Ähm, und dann halt immer, da hab' ich die Rohdaten durchgehen müssen und eben diese einzelnen Variablen raussuchen, was ich dann auch gemacht habe. Ähm ... die Rohdaten an sich waren selbst beschreibend, also was die Beschreibung der, also, welche ID des Patienten war das. [...]
37. **B:** Und dann habe ich das kontrolliert ob das jetzt stimmt z.B., ja. Aber das war ähm relativ zeitaufwendig, das Ganze. Also viele Instanzen. Man muss immer doppelt kontrollieren aber im Endeffekt waren alle Daten vor/ vorhanden. Das war schon mal der Vorteil.
38. **I:** Wie bist du denn darauf gekommen, dass es dort überhaupt solche Daten gibt, die

dich vielleicht interessieren könnten?

39. **B:** Aha, das war eine Mitforscherin, die gesagt hat, dass diese Studie durchgeführt worden ist und dass die Daten existieren und die passen jetzt zu den aktuellen Daten, ob man die nicht anschaut. Das war oral, mündlich überliefert. Ich glaub', das war jetzt nicht irgendwie offen einsehbar auf einem Server oder sowas, nein.
40. **I:** Verstehe.
41. **I:** In welcher Sprache werden denn die Forschungsdaten in deiner Disziplin beschrieben? Falls es sowas gibt, irgendeine Beschreibung.
42. **B:** Eine Mischung aus Englisch und Schwedisch.
43. **I:** Mm hmm, und die Gründe dafür?
44. **B:** Je nachdem, der erste der das eingibt, der bestimmt wie das dekodiert wird zum Beispiel. Ich hab' meistens auf Englisch kodiert, einfach weil es leichter war auch in der statistischen Auswertung dann und in der Präsentation, das ist dann eh auf Englisch, es wird nachher auf Englisch publiziert. Dann ist es halt leichter das gleich auf Englisch zu machen. Aber es war eine Gruppe, wie du gesagt hast. Das heißt, es sind immer wieder, sind halt schwedische Begriffe auch reingekommen.
45. **I:** Sind dir denn Metadatenstandards ein Begriff?
46. **B:** Nein.
47. **I:** Gibt es irgendwas in deiner Disziplin das du als standardisiert beschreiben würdest? Irgendeine Beschreibung der Daten, das standardisiert ist.
48. **B:** Eine. Also jetzt von Meta/ also von Daten, die ich produziert habe, oder die ich auswerte? [...]
49. **I:** OK, ich frage noch mal, ich weiß, das ist eine komplizierte Frage. Z.B. mit dieser Datenbank, von der du gesprochen hast. Die Daten dort müssen irgendwie beschrieben sein, sodass du weißt was es dort eigentlich für die Daten gibt. In welcher Form werden die Daten denn so dargestellt? Wie werden die beschrieben, sodass du weißt, was für Daten du da eigentlich hast.

50. **B:** Numerisch, tabellarisch. Also in einem Excel-File zum Beispiel oder in einer anderen Form von Datenbank.
51. **I:** Aber vor du auf diesen File zugreifen kannst, wie ist da der Vorgang? Ich nehme an, du kannst, das war vielleicht der Fall bei diesem [anderen] Krankenhaus, wo du direkt auf die Daten zugegriffen hast sozusagen dann. Aber bei dieser Datenbank, wie funktioniert das? Hast du da direkt überall die Excel-Tabellen und du kannst direkt auf die zugreifen? Oder ist da irgendein Zwischenschritt, wo du dir anschauen, was ist das eigentlich für Daten. Sagt dir das überhaupt irgendwas oder ist das komplett...?
52. **B:** Nein. Also, vielleicht verstehe ich's aber nicht. [...] Daten werden für mich produziert aus irgendwelchen Untersuchungen zum Beispiel jetzt, ja. Und wie gesagt. Und das wird dann irgendwo, meistens in externen Systemen, wie zum Beispiel diese Rohdaten meistens so gespeichert und überführt, ja, in tabellarischer Form halt ja. [...] Es gibt keinen einheitlichen Standard, weil jede Datenbank ist ein bisschen unterschiedlich. Also diesen Standard bestimmt man, also wie die Datenbank aufgebaut wird, bestimmt man eigentlich am Anfang der Studie. Man kann es auch bestimmen bevor irgendeine da irgendwelche Daten generiert werden, wie so 'ne Datenbank ausschauen wird und was da reinkommen soll und in welcher Form auch. Ganz logisch eigentlich, weil man hat eine Fragestellung. [...]
53. **I:** Verstehe. Welche Probleme hattest du möglicherweise bei der Beschreibung deiner Forschungsdaten? Oder der Forschungsdaten womit ihr im Team gearbeitet habt.
54. **B:** Reden wir jetzt immer noch über die Metadaten oder reden wir über meine Daten an sich?
55. **I:** Über die Metadaten.
56. **B:** Wir reden noch über die Metadaten, gut. Ähm, ja, das, was ich halt angesprochen hab'. Also es war teilweise nicht auffindbar, wo überhaupt Daten vorhanden sind. In welcher Form sie vorhanden sind, sind das jetzt wie gesagt Zettel oder ... Files irgendwo ähm. Und da muss man natürlich auch von einer Person wissen, dass es das überhaupt gibt, ja, eine Datenbank. Das heißt, das weiß man im Vornhinein

nicht.

57. **B:** Also man muss. Ah, mittlerweile ist es ja so, wenn du deine Daten publizierst, ja, dann musst du ja bekannt geben, wo sind diese Daten und wer ist dafür verantwortlich beziehungsweise mehr und mehr Journals gehen ja auch daraufhin, dass du deine Daten jetzt einfach öffentlich zugänglich machst. Anonymisiert. Das haben wir eh auch gemacht bei einer Studie, wo einfach die Daten, so wie sie sind, Datenbank, die wird publiziert, anonymisiert und jeder, der sich die Studie durchliest, kann theoretisch diese Daten nochmal analysieren und schauen, ob das stimmt, was wir analysiert haben. [...]
58. **B:** Das wird online publiziert, ja. Ich glaub, *PLOS One* ist nur online. Das ist nur eine online Publikation insgesamt. [...] Beziehungsweise, es gibt auch andere Journals, die ähm vorschlagen, dass die Daten auch auf einem externen Server gespeichert werden, die frei zugänglich sind. Mehr und mehr Journals machen das. Die Journals, die das nicht machen, die sagen halt, wo sind diese Daten und wen können Sie kont/, also wen können Forscher kontaktieren, damit man zum Beispiel Daten bekommt. Das ist ja die dritte Instanz. Es gibt hier auch die vierte, wo gar nichts irgendwo publiziert wird und keiner weiß, wo diese Daten sind.
59. **I:** Du hast gesagt, du hast schon so Daten von anderen auch verwendet, die andere erhoben haben. [...] Was wäre für dich wichtig, an dieser Metadatenebene? Bei der Beschreibung der Daten, sodass du die gut weiterverwenden könntest.
60. **B:** Ähm, relativ einfach sag' mal so. A, welche Studie das war? B, ähm wie war die Studie kodiert zum Beispiel, ja. Weil meistens ist ähm bei der statistischen Analyse, da hast du vielleicht, ähm, sind ja Daten geblindet. Also, wir wissen zum Beispiel nicht, das war bei uns auch so, weil am Anfang haben wir nicht gewusst, ob das Exposition war, dass jetzt Zigaretten oder nicht oder, ja. Das ist so analysiert worden, dass wir gewusst haben, es war Exposition A und B, ja. So wird das analysiert, dass sowas passiert bei A und was passiert bei B, ja. Und dann am Schluss, wenn die Daten fertig sind, die Analyse fertig ist, dann äh können wir sagen dann machen wir halt den Schlüssel auf und dann sehma, ob das A zum Beispiel Zigaretten oder nicht. Und das ist halt wichtig, dass das halt in so einer Datenbank, die schon fertig ist, und wo die Studie abgeschlossen worden ist, dass dieser Schlüssel auf jeden Fall vorhanden ist! Weil wenn der verschwindet, wissen wir's nicht im Nachhinein. Das

ist ein Kernpunkt, ja. Bei uns war das a Zettel. [...] Das hätt' ich auf jeden Fall irgendwo digital gesichert, ja. Aber, das kommt jetzt mehr und mehr, glaub' ich. Hoff' ich zumindest, ja. Gut, das ist mal ein, der wichtige Punkt, ja.

61. **B:** Dann zum Beispiel ähm allgemeine Info, wer hat diese Daten erhoben natürlich. Ähm, wann ist diese Studie gelaufen. Wo befinden sich überhaupt die Rohdaten, ja. [...] Diese Rohdaten müssen ja nämlich immer aufbewahrt werden für einige Zeit. Ähm und wo befindet sich die Datenbank an sich. Und dann von der Datenbank natürlich, eine, das ist dann relativ einfach, eine Beschreibung, also wie ist diese Datenbank aufgebaut ... und das reicht eigentlich. Also dass die Daten beschrieben werden und das halt innerhalb der Datenbank, wenn es da Codierungen gibt, dass die auch beschrieben werden.
62. **I:** Verstehe. Stell dir jetzt mal vor, dass nach dem Abschluss deines Forschungsprojektes jemand auf deine Daten zugreifen möchte. Wie wäre dies möglich?
63. **B:** Ja, das ist schwierig. Also, wie gesagt, die online publiziert sind, das ist ja ganz einfach, da bin ich aus dem Schneider. Da, wo ich eigentlich angegeben hab', dass ich die Daten besitze, dann müsste mich einer dieser, der die Publikation liest, mir ein E-Mail schreiben, ja. Und ähm mich fragen nach meinen Daten, ja. Aber das ist dann immer schwierig, weil wie weiß ich, dass das eine seriöse Person ist und diese Daten auch ähm verwenden will, also möchte ich jetzt [unv.], dass ich mit diesen Personen telefoniere und eigentlich ja, wer bist du? Warum möchtest du diese Daten? Wenn das halt dein Forschungskollege ist, da kann man halt dann sagen, also ich kann dir anonymisierte Daten zum Beispiel digital schicken. Das wäre eine Möglichkeit.
64. **B:** Und die Beschreibung von meinen Daten ähm ... wo ist das. Also die sind dann an sich kodiert in einer, Daten/ also zum Beispiel in SPSS sind die kodiert, ja. Ähm da müsst' ich ja aber dann das schicken, und dann weiß er, oder er sie, dass was das bedeutet, ja. Oder ich müsste dann noch dazu erklären, was das ist. Aber ich habe jetzt keine eigene File oder sowas, die, naja doch, es gibt ja Studienprotokolle. Man kann in Studien, die sind natürlich immer auf Englisch, und die kann man halt dazu fügen und dann kann sich diese Person da durchbeißen. [...]

65. **I:** Sind dir den Schulungs- oder Beratungsangebote bezüglich Forschungsdaten Management bekannt?
66. **B:** Ja. Also, [an der schwedischen Universität] gab's mehrmals ... so ein E-Mail-Ausschicker, wo eben dieses System, was ich vornhin erklärt hab' mit diesem Hotel, angepriesen wurde. Und es gab mehrere Info-Meetings dafür.
67. **I:** Hast du da teilgenommen?
68. **B:** Nein, hab' ich nicht. (lacht) Es war immer zu ungünstigen Zeiten und es ist alles andere wichtiger als das. Böse gesagt.
69. **I:** Verstehe. OK, jetzt kommen wir eigentlich schon zu meiner letzten Frage. Stell dir jetzt mal vor, dass alles möglich ist. Finanzen spielen keine Rolle. Welche Services oder Unterstützung seitens der Universität könntest du gebrauchen bei der Erstellung von Metadaten für deine Forschungsdaten oder generell beim Forschungsdatenmanagement?
70. **B:** Mm, also, das wäre vielleicht ein Team von Leuten eigentlich, ja. Ähm die verschiedene Kompetenzen haben. Also da würd' ich sagen, einer oder eine ist spezialisiert auf was das Rechtliche angeht. Das ist sehr sehr wichtig, find' ich. Ähm aus mehreren Gründen einfach. Also weil man muss höllisch aufpassen, dass die Ethik stimmt, dass nichts gebrochen wird von der Ethik. Und das hat auch Konsequenzen auch für die Publikation, dass das alles wirklich eingehalten wird, ja. Das ist vielleicht eine Person, die sich da eben auskennt, also national. Also wie das wird, das ist von Land zu Land verschieden oder von teilweise vielleicht sogar Uni zu Uni. Aber eine Person die sich damit auskennt.
71. **B:** Dann halt eine weitere, die sich halt statistisch oder mathematisch oder so auskennt, ja, wie man sowas am Besten anlegt. Weil das ist ja irgendwie eine Datenbank über Daten, ja. Und da ist meistens so ein, also ich weiß nicht, ob das ein Statistiker ist oder gibt's da irgendwen, der da speziell auf Datenbanken sich spezialisiert. Aber in diese Richtung meine ich jetzt, ja. Und äh vielleicht administrative Personen, ... die dann so Sachen ausführen können, was die anderen eintragen, ja. Ähm, wer noch? ... Ja, und dann also, so ein Team, was zur Verfügung gestellt wird zum Beispiel der gesamten [Universität], hmm. Und da schickt man ja eben ein, also wir haben so eine, wir planen so und so eine Datenbank, ja. Können

Sie sich das anschauen, also ob das jetzt konform ist mit der Rechtslage und wie das jetzt technisch durchführbar ist. Und dann geben die das weiter an Beschäftigte dort. Und sagen, gibt, helft der Gruppe, baut eine Metadatenbank auf. Man braucht nicht mehr ganz viel Geld dafür, würd' ich sagen, aber a bissel was auf jeden Fall. Und ich weiß nicht wie groß so ein Team sein sollte, damit das halt so viele Forschergruppen abdeckt. Schwer zu sagen.

72. **I:** OK, hast du jetzt noch irgendwas zu sagen zu diesem Thema, was ich vielleicht nicht in den Fragen abgedeckt hab'?

73. **B:** Nein, das war 'ne interessante Fragestellung. Also, ich hab' nicht wirklich darüber nachgedacht aber jetzt im Nachhinein natürlich, dass wo ich mal da so zurückdenke, das mit diesen Papieren und Büchlein, also dass das eigentlich eine relativ ganz gute Idee ist, das von Anfang an richtig zu machen nämlich. [...]