

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**

**(unv.) = unverständlich / unintelligible**

1. **Wissenschaftler A – Teiltranskript, persönliches Interview (24.04.2019)**
2. **I:** OK, ich weiß zwar schon ein bisschen über dich aber wenn du mir einfach zum Anfang dein Forschungsgebiet beschreiben würdest.
3. **B:** Also kurz zu mir, ich habe [letztes Jahr mein Medizinstudium abgeschlossen und befinde mich im ersten Arbeitsjahr an einer Universitätsklinik in Wien]. Und ähm ich befasse mich konkret mit autoimmunen Muskelentzündungen. Das ist eine Krankheitsgruppe, die sehr selten ist. Ähm, sch/ unter Umständen auch schwer zu behandeln, aber es tut sich vieles. [...] Das interessante ist, ja halt (unv.), dass diese Entzündung nicht nur an die Muskulatur beschränkt ist, sondern auch auf die inneren Organe übergreifen kann. Insbesondere, was auch Prognose entscheidend ist, ist der Lungenbefall und ich bin jetzt da, quasi der mehr oder weniger, der Junge hier im Haus, der die Aufgabe erteilt bekommen hat, alle diese Patienten dieser Kohorte des [Krankenhauses] aufzuarbeiten, aufzuholen. Und ähm das mach ich auch. Ich bin ähm, mein Bestreb ist jetzt eine Datenbank zu generieren, diese Patienten systematisch auf/ zu erfassen und natürlich auch, ja, dahingehend auch Publikationen zu machen.
4. **I:** OK, super. Genau, kannst du mir jetzt bitte kurz definieren, was in deiner, eurer Disziplin als Forschungsdaten gilt?
5. **B:** Als Forschungsdaten jetzt konkret gelten natürlich personenbezogene Angaben wie Alter, Geschlecht, und auch die ethnische Zugehörigkeit. Weil wir wissen, dass bei Autoimmunerkrankungen, zum Beispiel beim Lupus, gibt es eine, ein enthäufteres Vorkommen bei Frauen asiatischer Raum. Diese natürlich, dann natürlich auch die Familien, familiäre Gegebenheiten, gibt es sozusagen eine familiäre Häufung von Autoimmunerkrankungen. Also quasi die personenbezogene, patientenbezogene, serologische, also Blutwerte, Labor, da schauen wir uns ähm gewisse, ich glaub', es wird zu viel wenn ich jetzt ins Detail gehe welche. [...] Und natürlich auch ähm patientenbezogene Angaben, nämlich

Assessments, patient-reported Outcomes. Also anhand dieser Parameter, beziehungsweise Domänen versuchen wir da die Daten, aus diesen Domänen, die Daten zu generieren.

6. **I:** Alles klar, OK. Sagt dir der Begriff Metadaten irgendetwas?
7. **B:** Ähm, Metadaten? Ähm.
8. **I:** Genau, was stellst du dir darunter vor?
9. **B:** Ich weiß, was das Wort, ich kenne Metaanalysen, aber ich, was Metadaten jetzt sind, Definition kann ich jetzt nicht so bringen.
10. **I:** Was würdest du dir da so vorstellen, unter dem Begriff, wenn du das hörst?
11. **B:** [...] Ich würde eher meinen, dass das ähm ... Ich muss gestehen, ich meine ja gut, mit dem Wort Daten kann ich was anfangen und mit dem Meta auch. Und jetzt zusammen zu schließen ja ... so übergeordnete Datensammlung, irgendwie so, in diese Richtung. Das wär's ja.
12. **I:** [...] Ich hab' hier so ne', eine von den vielen Definitionen, kannst du dir mal anschauen.
13. [Lesen von Definition]
14. **I:** Jetzt ähm wäre meine Frage, wenn du jetzt diese Definition gesehen hast und wenn wir das ein bisschen klarer gemacht haben ähm. Wenn du selber wissenschaftlich tätig bist, erstellst du sowas, solche Metadaten, solche Beschreibungen für deine Daten?
15. **B:** Ja, natürlich. Für die Systematik ist der, es scheinen die Metadaten ein essentielles Tool zu sein, oder. [...]
16. **I:** Genau, und wie beschreibst du denn deine Daten in deinem speziellen Fall? Wie funktioniert das, wie kann ich mir das vorstellen?
17. **B:** Konkret jetzt, also diese Krankheit, diese Muskelentzündungen, wenn ich das jetzt anhand dieses Bespiels oder meines Falles konkretisieren darf' ähm. Ähm unterteilt man in einzelne Subgruppen und die sind natürlich untereinander sehr heterogen, sprich ich tu' mir um das zum Vereinfachen zum nächsten mal beschreibend. Eben wie viele Patienten ich hab', sprich numerische Daten und auch dichotom, also männlich weiblich. Dann natürlich auch numerische Variablen, nämlich Laborparameter und auch ... wie ich gesagt hab' Bildgebung, Lungenbefall, ja nein, Lungebefall welches Muster. Also es ist zunächstmal beschreibend. So würd' ich jetzt a mal diese Datenbank jetzt a mal definieren. In Therapiestudien würd' ich Therapie, einzelne Therapiestudien planen mit dem Patienten, wäre es nicht viel anders. Sprich, ich beschreibe dann nur, wie sich die

einzelnen Parameter verändern anhand dieser Therapie. Und versuchen anhand einem gewissen Score, der international anerkannt, also standardisiert ist, zu objektivieren [...]. Also zusammengefasst eher beschreibend und vergleichend.

18. **I:** Kannst du mir bitte jetzt schildern, wozu diese Metadaten, diese Beschreibungen dienen? Jetzt nicht die Daten selbst, aber wirklich die Beschreibung von denen.
19. **B:** Ähm, meinst du was der Sinn von dem ist?
20. **I:** Genau.
21. **B:** Von dem Sinn A, wir haben eine systematische Erfassung von allen diesen Patienten mit dieser seltenen Erkrankung, die hier in [der Stadt] sind. Mein', wir hätten sozusagen jetzt mal unizentrisch. Die Idee ist, das zunächst a mal auf multidisziplinär zu übergreifen, sprich, wir sind die Rheumatologie. Diese Krankheit hat auch einen Hautbefall, gut, dann würden wir die Dermatologen mit ins Boot holen. Diese Krankheit hat auch einen Muskelbefall und der irgendwie muss ja jemand diesen Muskel beschreiben, das sind die Pathologen. Dann gibt es auch einen neurologischen Ansatz, die sind auch die, die daran interessiert sind. Und für die ganze Veränderungen in der CT zum Beispiel der Lunge die Radiologen. Also es wäre ein multidisziplinärer Ansatz. Das wäre jetzt in dem Hause beschränkt. Die Idee ist es natürlich noch weiter das auszubauen. Sprich, wir würden dann daraus multidisziplinär und in weiterer Folge multi/ ähm multizentrisch werden, also mehrere Zentren. [...]
22. **I:** In welcher Sprache erstellst du denn diese Metadaten, diese Beschreibung von den Daten?
23. **B:** Ähm, beides, bilingual. Deutsch und Englisch eigentlich.
24. **I:** Und was sind da so die Gründe dafür? Was würdest du so sagen? [...]
25. **B:** Deutsch natürlich. Weil A, das sozusagen eine Austriadatabank, oder? Ähm, natürlich kann man, und um international ... natürlich jetzt a gewisse, wie soll ich sagen, um dem internationalen Rahmen zu genügen ist natürlich jetzt absolut die englische Sprache indiziert. Und jetzt wie hier, in unserem Jargon, wir brauchen doch ab und zu englische Vokabel, Termini [...] Ja, es ist eine gute Frage. Ich kann's auch ganz eigentlich nicht be/ absolut begründen aber es wäre. Deutsch ist natürlich jetzt a mal für diese Ö/, weil Österreich und Englisch ist doch, weil ich doch ähm in die, um eine, was soll ich sagen, um eine, mit dem Ziel, irgendwas zu veröffentlichen. Quasi um den internationalen Normen zu entsprechen aber. Ich würde jetzt mal meinen, primär Deutsch und in weiterer Folge dann natürlich auch Englisch, ja. Reicht dir das als Antwort oder kann man das nachvollziehen, oder?
26. **I:** Ja, klar. [...] Du hast schon ein bisschen jetzt von Normen gesprochen und von

Systematik und, genau. Sind dir denn, wendest auch Metadatenstandards [...] an? Oder ist es dir bekannt?

27. **B:** [...] Aber, nur kurz, ich muss ganz genau verstehen, was du damit, mit der Frage meinst? [...] Es ist standardisiert.
28. **I:** Alles?
29. **B:** Ja, nicht, alles nicht, alles. Sondern, ich kann jetzt auch. Weil ... ich ähm, das komplizierte ähm, oder das, bei den rare diseases ist es, dass es nunmal. Es gibt schon mehr oder weniger standardisierte aber nicht absolut validierte Sachen und deswegen hat man doch einen gewissen Freiraum, in dem man sich da bewegen darf. Ein Beispiel, nur ein konkretes Beispiel. Ein Patient, mit dieser Muskelerkrankung [...] Es gibt keinen richtigen, fassbaren oder absolut standardisierten Tools, wie ich diese Atemnot produktiviere. Und das erlaub' ich mir schon. Dann verwende ich halt aus meinem sozusagen, ich kann schon ein paar Tools anwenden, wo ich zumindest probieren kann, um diese Variable Atemnot zu assessen. Also, es gibt schon ein paar, um einen sozusagen internationalen Vergleich zu machen, muss ich mich doch an die Regeln halten. Ich muss an diesen Scores festhalten, dass ich meine, dass ich äh sozusagen die Power meiner Studie erhöhe. Dazu hab ich die. Aber ich kann auch (unv.) was innovatives machen. Das bietet mir schon. [...] Also, ich bin nicht absolut gefangen. Sondern kann auch, aha, das wurde noch nicht erfasst, dann probieren wir's mal. [...] Also gewissen Freiraum hammer schon, aber das ist vielleicht, was man bei Rare Diseases mehr hat.
30. **I:** Alles klar. Hattest du denn schon Probleme bei der Beschreibung deiner Daten?
31. **B:** Die Probleme gibt's immer. [...] Das ist halt so, dass nehmen wir, ein Patient bekommt diese Erkrankung. Dann ist es so, wenn ich eine Therapie starte, wäre es im Idealfall, ich unterstreiche das Wort Idealfall, dass ich den Patienten in regelmäßigen Follow-up-Untersuchungen neuobjektiviere unter einer gewissen Therapie. [...] Was ich sagen will, die Follow-up-Untersuchungen finden nicht immer regelmäßig statt. Es wird ausgefüllt, es wird vergessen. [...] Also, das da, also die regelmäßigen Follow-up-Untersuchungen. Und wenn ja, wurden überhaupt alle gemacht? Damit ich halt eben doch diese, dass ich diese internationalen Scores ermitteln kann, hab ich nicht immer manchmal Variablen, um es zu machen beziehungsweise kaum. [...]
32. **I:** Kannst du mir bitte beschreiben, welche Rolle die Metadaten, die Beschreibung deiner Daten spielt, wenn du im Team arbeitest? [...]
33. **B:** Das muss natürlich sein, ja. Also spielt natürlich eine Rolle, weil ähm, man erkennt eben Muster. Man erkennt Verknüpfungen, Assoziationen [...] Ja sicher, naja, eben auch deswegen, weil diese Krankheit auch multidisziplinär behandelt wird, natürlich wäre es natürlich interessant wenn ich vom Dermatologen erfahre, wie viele Patienten hat er überhaupt, die ich nicht kenne. Und wenn er

sagt, ich habe so und so viele Patienten und ich bin eigentlich mit der Therapie sehr gut gefahren dann, oder? [...]

34. **I:** Hast du denn schon mal auf Forschungsdaten von anderen zugegriffen? Hast du Forschungsdaten von anderen benutzt?
35. **B:** Absolut, ich bin davon abhängig.
36. **I:** Genau. Super. Und, was hattest du da denn für Erfahrungen mit den Metadaten von den Anderen? Mit den Beschreibungen, wie die Anderen deren Daten beschrieben haben?
37. **I:** Vorbildliche, weil ich weiß, die machen es. Aha, die haben das so gemacht, gute Idee. Also ich komme auf Ideen. Und ähm Patientenzahl ist, wie gesagt, rare Disease, niedrige Zahl. Ich bin a mal auf deren Input angewiesen aber auch auf deren Kooperation, weil nur so kann man, glaub' ich, die Power erhöhen. Und, aber ich hab da Ideen bekommen und so konnt' ich auch zum Beispiel erfahren, die haben es so gemacht. Ich hätte das zum Beispiel anders gemacht und auch komplett andere ... Ideen für mögliche retrospektive Untersuchungen, prospektive Untersuchungen, also sprich durch dieses Studium oder durch das Lesen von anderen Daten und Datenbanken hab' ich nicht. Hab ich natürlich A, Ideen. B, schon wissenschaftliche Erkenntnisse und die versuch' ich auch an meinen Patienten dann anzuwenden. Wenn ich weiß, das es sogar Therapien und so weiter, die und die besser ansprechen, das heißt, die haben ja schon irgendwas in der Richtung schon gemacht. Hab' ich neue Therapieansätze ... und ja.
38. **I:** Wie kann ich mir das so vorstellen, wenn du so Daten von Anderen, wenn du auf die zugreifen möchtest, wie funktioniert das? [...] Wie findest du diese Daten?
39. **B:** Wir haben da unsere ... Datenbanken online, diese PubMed ... wo man Publikationen herausnimmt. Und, die einfach sozusagen studiert. [...]
40. **I:** Und wenn du dir jetzt mal vorstellst, dass nach dem Abschluss deines Forschungsprojektes [...], dass eine andere Person auf diese Daten zugreifen möchte, wie wäre dies möglich?
41. **B:** Es wäre möglich, weil man (unv.) zum Beispiel, nehmen wir mal, reden wir von Österreich. Es wäre insofern, weil wir dadurch ein Data sharing machen würden. Wie ich diese Data sharee, hat einen (unv.) Vorteil. Der Vorteil ist zumal der, dass wenn er mit meinen Data was macht, bekomme ich eine Ko-Autorenschaft, es ist ja quasi auch das Interesse zunehmend natürlich nur so Gelder hineinkommen, in die Abteilung. Also ich hätte mal dieses Sharing ko/ unter Daten bringt Publikationen und auch eine Kooperation. Ähm, ich sehe bei rare Diseases das Data sharing als vorteilhaft und es ähm gibt doch die Idee, beziehungsweise es existiert auch schon eine internationale Imucities. Imucities ist diese

autoimmune Muskelentzündungen Datenbank, sprich, wenn man das beisteuert, wenn man Patientendaten beisteuert. Ähm wird man natürlich in dem Clus/ in den Empfehlungen, in den Kriterien der Diagnose miterwähnt. Und das ist natürlich für eine Abteilung erstrebenswert. Also, ich sehe das bei rare Diseases nochmal deshalb eine Data sharing nur vom Vorteil.

42. **I:** Das heißt, wenn du jetzt mit dieser Datenbank fertig, was machst du damit, um dieses Data sharing zu gewährleisten?
43. **B:** Ja also, kommt mmmm. Nachdem die, natürlich, die ethischen Voraussetzungen geklärt sind, ähm, wir kooperieren. Kooperation. ... Ich biete meine Daten an, bekomme dafür eine Ko-Autorenschaft. Weil ich natürlich im Gegenzug auch ein Projekt plane, dann, um die Power zu erhöhen würd' ich denn natürlich nochmal zu dir zurückzukommen und sagen: Hey, kannst du mir bitte deine Daten borgen, ich habe etwas geplant. Und so ist es ... Also, so in etwa, so würde das laufen. Es ist, wie gesagt, vielleicht bin etwas gebiased oder von rare Diseases aus, sagt mal so. Wir sind voneinander mehr oder weniger abhängig und von Kooperation abhängig. Was ich jetzt durchaus als, muss ich gestehen, als positives sehe.
44. **I:** Sind dir denn so Schulungs-, Beratungsangebote im Bereich Forschungsdatenmanagement, sprich auch Data Sharing zum Beispiel, bekannt?
45. **B:** Muss gestehen, nein. Ich lerne alles selber, beziehungsweise tu' ich auf Empfehlungen und Erfahrungen der anderen Kollegen, ja.
46. **I:** Und jetzt meine letzte Frage. Wenn alles möglich wäre, welche Services, welche Unterstützung würdest dir den wünschen vonseiten der Universität bei zum Beispiel der Erstellung der Metadaten für deine Daten, dieser Beschreibung?
47. **B:** Hmmm natürlich a mal infrastrukturell. Infrastrukturell, sprich, ja, die Infrastrukturen seh' ich in einem Dienstzimmer mit einem Computer, der [...] fähig ist, schnell arbeitet, das für Infrastruktur. Vielleicht auch an Personal, um das natürlich zu beschleunigen, diese Aufarbeitung der Patienten und ähm Gelder beziehungsweise für gewisse Biopsien, mein', da gibt's spezielle Untersuchungen, die du machen musst, da braucht man natürlich für alles Geld. Und ja.
48. **I:** Wenn ich noch kurz nachfragen darf, bei dem Personal. Was meinst du genau?
49. **B:** Zum Beispiel Medizinstudierende, die eine Diplomarbeit benötigen.
50. [...]