

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 H – Teiltranskript, persönliches Interview (11.05.2019)**
2. **I:** Okay, kannst du dich bitte zum Anfang kurz vorstellen und dein Forschungsgebiet beschreiben.
3. **B:** Ja, sehr gerne. Mein Name ist Wissenschaftler H. Ich bin [Arzt und Junior Researcher an einer Universität in Wien] [...] Meine Forschungstätigkeit ähm beläuft sich im Moment ähm zum einen aus der Forschungstätigkeit, die noch im Rahmen meines, noch im Zusammenhang steht mit meinem Studium, meiner Doktorarbeit, die ich an [einer anderen Universität] verfasst hab'. Eine retrospektive Studie zu nicht kleinzelligem Bronchialkarzinom bei älteren Patienten. Aber mein aktueller Forschungsschwerpunkt ist die interstitielle Lungenerkrankung und da insbesondere ähm der Aufbau eines ähm Registers, eines prospektiven Patientenregisters.
4. **I:** Sehr schön. Kannst du jetzt bitte definieren, was in deiner Disziplin als Forschungsdaten gilt?
5. **B:** Also, in meiner Disziplin, da würd' ich jetzt mal sagen in der Medizin an sich, ähm sind Forschungsdaten sehr ähm mannigfaltig und unterschiedlich. Ich würde mal sagen, es gibt zum einen objektivierbare Patientendaten, sowas wie Laborparameter. Oder bei bildgebenden Verfahren die Größe und die Ausdehnung von Organen ähm oder gerade in der Pulmologie, der Lungenheilkunde ähm

lungenfunktionelle Verfahren, die Forschungsdaten generieren, sei das Lungenvolumina ähm oder auch dynamische Volumina. Und darüber hinaus gibt es natürlich den ganzen Aspekt der Patienten-zentrierten Daten. Das sind dann Daten, die jetzt nicht rein objektivierbar sind, sondern die auch auf den Patienten-Input angewiesen sind. Ähm, Fragebögen, Questionnaires, die halt nicht nur objektivierbare Daten ähm verwenden, sondern auch vom Patienten empfundene Symptome, sei's Atemnot oder Schmerzsymptome oder andere Aspekte.

6. **I:** OK, sagt dir der Begriff Metadaten irgendetwas?
7. **B:** Ja, hab' ich schon mal von gehört. Ja.
8. **I:** Was stellst du dir darunter vor?
9. **B:** Unter Metadaten stell' ich mir vor ähm. Ja, letztendlich, vom griechischen Ursprung des Wortes her, Meta, als über etwas. Also, irgendwie Daten die Daten beschreiben, sei es jetzt irgendwie zum Beispiel ähm Datumsangaben oder Datenformate. Wie verschiedene Datensätze letztendlich elektronisch sozusagen, was für Formate die gespeichert sind. Oder ... ähm auch Zugriffsrechte vielleicht, vielleicht sind es auch Metadaten, würd' ich sagen. Also alle möglichen Daten, die über, ja, über Datensätze sozusagen einsehbar sind. Erstellungsdatum, Zugriffsrechte, Datenformate und so weiter und so fort.
10. **I:** Sehr schön. Wenn du selber wissenschaftlich tätig bist, was für Metadaten erstellst du für deine Forschungsdaten?
11. **B:** Hmm, da muss ich kurz überlegen. Ähm, ja, was für mich immer wichtig ist, sind Daten zur oder Metadaten bezüglich der Erstellung und des Zeitpunkts und der Bearbeitung ... von Daten. Also das ist einfach der Chronologie-wegen, dass ich auch immer nachvollziehen kann, wann ich was gemacht habe. Ähm, dass ich halt irgendwie kein Logbuch führe oder sowas, sondern über die Metadaten vor allem über die Erstellungsdaten und Bearbeitungsdaten nachvollziehen kann, wann was bearbeitet wurde. Was auf dem aktuellsten Stand ist und was nicht, ja. Das wäre für mich wichtig, ja, Daten zur Erstellung. Teilweise sind die Daten dann auch Passwort geschützt, aber ich weiß nicht ob das jetzt Metadaten sind. Das ist sicher, vielleicht wieder ein anderer Aspekt. Weiß ich nicht. Ja.

12. **I:** Kannst du bitte kurz schildern, wozu diese Metadaten dienen?
13. **B:** Ja, wie jetzt gerade schon angesprochen, insbesondere der, ähm, das Wiederfindens von Daten, ja. Und auch des Erkennens der Aktualität sozusagen von Daten. Ähm, also dafür, würde ich sagen, sind Metadaten sehr hilfreich, ja. Das ist aber auch so ziemlich, glaub' ich, der vordergründigste Aspekt von Metadaten für mich. An dieser Stelle.
14. **I:** Kannst du bitte beschreiben, welche Rolle Metadaten spielen, wenn du im Team arbeitest? Zum Beispiel anhand deines letzten oder jetzigen Forschungsprojektes.
15. **B:** Hmm. Ich muss gestehen, dass meine Forschungstätigkeit, dass ich relativ eigenständig und solitär irgendwie arbeite. Und ich Daten selbst gar nicht so viel teile, weniger als jetzt schon die Auswertungen oder die Analysen von Daten oder schon, weiß ich nicht, Artikel oder so. Ähm, deshalb würd' ich sagen das Teilen von Daten, ja. Mach ich gar nicht so viel. Und wen natürlich, gut, ja. Was war die Frage nochmal?
16. **I:** Kannst du bitte beschreiben welche Rolle Metadaten spielen, wenn du im Team arbeitest.
17. **B:** Gut, ähm ... Metadaten im Team. Ja, vielleicht auch Zugriffsrechte irgendwie. Dass das immer wichtig war, wer Zugriff hat auf was. Ähm, dass man ja, gerade in der Medizin hat man ja oft Rohdaten von Patienten, die sind dann oftmals nicht anonymisiert in einer Form der Pseudonymisierung oder vollständigen Anonymisierung. Sodass man halt den Namen und das Geburtsdatum und andere personenstandsrelevanten Daten. Und um diese Daten dann weiter verwenden zu können, oder auch teilen zu können, datenschutzrechtlich, datenschutzkonform, muss man diese dann pseudonymisieren. Ähm das ist sicherlich ein Aspekt, der beim Data Sharing immer eine große Rolle spielt, dass man halt die Daten pseudonymisiert, vor man sie teilt, ja. Aber das ist jetzt eigentlich auch wieder kein Aspekt von Metadaten, würd' ich sagen. Sondern einfach die Art der Daten, ja, ob die jetzt pseudonymisiert sind oder nicht, ja. Das hat immer 'ne wichtige Rolle gespielt. Sonst fällt mir dazu jetzt nicht viel mehr ein.
18. **I:** Verwendest du auch Forschungsdaten von anderen? Die von anderen erhoben wurden oder erzeugt wurden?

19. **B:** Ja auf jeden Fall, im Rahmen meines letzten wissenschaftlichen Projektes hab' ich ähm Forschungsdaten verwendet aus einem großen Krebsregister [einer deutschsprachigen Universität], die Patienten mit einem Lungenkarzinom erfasst. Und ich habe diese Daten verwandt, um eine retrospektive Datenanalyse ähm zu verwenden. Da hab' ich tatsächlich bestehende Daten ähm zur Einsicht und zur Verarbeitung erhalten, ja.
20. **I:** OK. Zum Beispiel, anhand dieses Beispiels. Welche Erfahrungen hast du mit den Metadaten von diesen Datensätzen oder von diesem Datensatz gemacht?
21. **B:** Ähm. Mit dem Datensatz selbst hatte ich so meine Probleme. Ja, gut. Das ist vielleicht ein Aspekt, stimmt, an den hab' ich vorher gar nicht gedacht. Ähm, die Daten waren zum einen nicht ganz vollständig und es haben auch einzelne Daten gefehlt, die dann noch zusammengesammelt werden mussten. Aber insbesondere bei der Klassifizierung von bestimmten Patientenmerkmalen, wo man sich dann halt eines numerischen Systems bedient hat mit eins, zwei, drei, vier, fünf oder eines anderen Systems war ... hat eine Legende gefehlt. Ein Hinweis darauf, wie ich sag' jetzt mal, zum Beispiel die adjuvante Therapie bei ähm Lungenkarzinom Patienten klassifiziert wird. Die Daten waren dann im Datensatz nur codiert als eins, zwei, drei, vier, fünf und ich wusste jetzt nicht welcher Patient hat eine Chemotherapie bekommen. Welche Patient hat eine Chemotherapie vor der Operation bekommen oder nach der Operation, welcher Patient hat eine Strahlentherapie bekommen. Welche Patienten haben womöglich beides bekommen. Also wenn man das zu Metadaten zählen könnte, dann die Erläuterung, Klassifizierung, Legende von Daten. Von numerischen Daten, wo eigentlich noch andere Sachen dahinterstecken, ja. Das war sicherlich unvollständig und ähm oder sagen wir mal ausbaufähig, ja. Da hatte ich dann meine Mühe ... damit.
22. **I:** In welche Sprache beschreibst du denn deine Forschungsdaten?
23. **B:** Also, auf Deutsch und auf Englisch.
24. **I:** Und was sind so die Gründe dafür?
25. **B:** Ähm, meine Doktorarbeit hab' ich auf Deutsch geschrieben, weil das vonseiten der Fakultät auch insbesondere so gewünscht wurde. Ähm und alle weiteren

Projekte bereits und noch zu publizierende Paper, da ist halt die Wissenschaftssprache die englische Sprache. Und ähm, dass ist die Sprache in der, ja, Wissenschaft stattfindet in der Medizin.

26. **I:** Sind dir denn Metadatenstandards bekannt? Und wendest welche an?

27. **B:** Muss ich verneinen. Nein. Also, nee.

28. **I:** Kannst du mir bitte erzählen, welche Probleme du möglicherweise hattest, bei der Beschreibung deiner Forschungsdaten? Bei der Erstellung von Metadaten.

29. **B:** Bei der Beschreibung meiner Forschungsdaten, bei der Erstellung von Metadaten. ... Gut ... muss ich kurz überlegen. Also bei der Beschreibung der Forschungsdaten hatte ich natürlich das Problem der Unvollständigkeit teilweise, dass gewisse Einträge gefehlt haben. Das hat mir ähm Probleme bereitet hier und da. Die Frage wie man fehlende Daten ergänzen kann. Ob es irgendwie Modelle gibt, fehlende Daten zu generieren oder ob man dann einzelne Cases ausschließen muss oder sollte. Ähm, jetzt insbesondere Probleme bei der Beschreibung von Metadaten ähm wären mir eigentlich nicht aufgefallen.

30. **I:** Du hast das schon kurz angeschnitten, dass du auch Forschungsdaten von Anderen verwendet hast. Was war dir, oder ist dir dabei wichtig, bei den Metadaten, wenn du Forschungsdaten von anderen verwendest?

31. **B:** Ja, 'ne gewisse Vollständigkeit zum einen, dass Daten verständlich sind. Überhaupt teilbar und reproduzierbar in einer Form, dass man sie versteht. Ähm und auch von den Formaten her zugänglich, ja. Dass man halt ähnliche Formate und Softwares benutzt. Ich hatte ähm letztendlich bei der wissenschaftlichen Arbeit, wo mir externe Daten zur Verfügung standen, hab' ich diese in einem Excel Format bekommen. Und hab' meine, die Daten dann in ein sozusagen eine Art Register eingespielt, dass ich mit dem Programm Filemaker erstellt habe.

32. **B:** Da ist natürlich Excel jetzt ein absolutes Standardformat, dass alle bedienen und können und kennen. Aber sicherlich gibt es da auch andere Formate, wie zum Beispiel die Filemaker-Formate, die ich dann auch mit meinen Kollegen geteilt hab'. Und manche haben da nicht die Software für, um das abspielen zu können. Also sicherlich der Aspekt der Reproduzierbarkeit, der Übersichtlichkeit, ähm des

Verständnisses der Daten. Dass diese auf einer Art und Weise zusammengetragen werden, dass man das versteht. Und auch von den Datenformaten her, dass man das re/, ja, einlesen, einsehen kann und auch bearbeiten kann, verwenden kann, ja. Das sind sicher Aspekte, die ich bei der, bei Metadaten wichtig finde, ja.

33. **I:** Stell dir jetzt bitte mal vor, dass nach dem Abschluss deines jetzigen Forschungsprojektes eine andere Person auf deine Daten zugreifen möchte und diese nutzen möchte. Wie wäre dies möglich?
34. **B:** Ja, das halte ich für ein sehr wichtiges und relevantes Szenario, weil ich ja. Was ich jetzt mache, ein Register, monozentrisch jetzt [an der Universität] beginne und im Idealfall auch dann in andere Institutionen weitertragen möchte oder in Kooperation mit anderen. Das heißt, der Aspekt der, des Data Sharings wird dann ja sehr wichtig sein. Dass man die ähm Parameter, die man erheben möchte, dass dann auch anderen mitteilt und idealerweise bestehende Datenbanken weiter bestückt und weiter bespielt. Und die Frage war, wenn ich richtig verstanden hab', welcher Aspekt beim Teilen dann für mich sehr wichtig wäre, richtig?
35. **I:** Wie wäre das möglich, dass jemand anderer auf deine Daten zugreifen könnte?
36. **B:** Wie das möglich wäre. Ja, genau. Ähm, da ist genau, dass ich halt indem ich Datenbanken verwende, die verständlich sind, einsehbar sind und die man auch teilen kann, genau. Und das ich halt natürlich auch datenschutzrechtlich, konform. Patientendaten in anonymisierter oder pseudonymisierter Form erhalte und diese auch teilen darf. Entsprechend des Informed Consent.
37. **I:** Wie kann ich mir das Data Sharing oder das Teilen von den Daten [...] vorstellen?
38. **B:** Das ist eine gute Frage, weil ich so viel Data Sharing noch nicht betrieben habe. In der Theorie würd' ich mir das dann so vorstellen, dass man ähm Zugriffsrechte zu den Daten generiert, die auf den Servern der [eigenen Universität] liegen. Wir arbeiten da auch mit der IT zusammen, dass man eine Datenbank benutzt, die auch in anderen Bereichen schon für Forschungsdaten verwendet wird. Dass man praktisch über verschiedene ähm künstliche Intelligenz-Programme ähm Patientendaten einspielt und auch gewisse Aspekte aus Befunden automatisch rauslesen und reinkopieren. Laborbefunde hineinkopieren kann, um schnell und praktisch Cases zu generieren. Ähm und die sind dann über die Server der

Universität einsehbar.

39. **B:** Da definiert man natürlich dann auch Zugriffsrechte, wer darf das einsehen, wer darf das nur einsehen, wer das/ darf das auch einsehen und bearbeiten. Und das Data Sharing wäre dann, dass man ähm diese Zugriffsrechte erweitert und andere Partner sozusagen mit ans Bord holt und die dann auch Zugriff haben auf die bestehenden, die bestehende Plattform, auf das bestehende Register. Und dieses dann mit eigenen Daten erweitern, indem sie neue Patienten einspielen oder neue Nachbeobachtungsdaten bestehender Patienten ergänzen.
40. **I:** OK. Sind dir denn Schulungs- oder Beratungsangebote bezüglich Forschungsdatenmanagement bekannt?
41. **B:** Es gibt ähm eine Stelle für Forschungsservice an meiner Universität. Ähm im Rahmen der Erstellung eines Ethikantrags, eines Ethikprotokolls zur Vorstellung in der hiesigen Ethikkommission, hab' ich einen Forschungsdatenmanagementplan erstellt und habe eine Forschungsdatenmanagementplan-Vorlage vonseiten meiner Fakultät eingefordert. Ähm, das kommt wohl regulär bei monozentrischen Registern eigentlich noch nicht so sehr zum Einsatz, aber mir wurde ... dann eine Vorlage für Forschungsdatenmanagementpläne von [einer anderen Universität] angeboten. Eine spezielle Vorlage für die Medizin/ also für [meine Institution] gab es in dem Sinne nicht. Ähm und das war halt ein relativ ausführliches Dokument ähm für Forschungsdatenmanagementpläne. Ähm, dass sicherlich dann noch doch recht weit weg war von dem, was ich brauche. Aber zumindest mal gewisse Aspekte des Forschungsdatenmanagements berücksichtigt hat von, ja., überhaupt was. Welche Daten ich erhebe und wie ich die ähm speichere, wie ich die archiviere, wie ich datenschutzrechtlichkonform arbeite. Und so weiter, ja.
42. **B:** Aber der Datenmanagementplan war sicherlich ähm noch relativ weit weg von meiner Realität. Für mein Empfinden etwas zu umfangreich und zu vage ähm formuliert. Ich würde mir wünschen, dass es vielleicht vonseiten meiner Fakultät für derartige Projekte dann noch Vorlagen gibt. Zum Beispiel dann auch direkt explizit, dass man unterscheidet zwischen klinischer und präklinischer Forschung, weil das ja doch teilweise sehr unterschiedlich ist. Ähm, ob man jetzt nur irgendwelche objektiven Parameter aus einem Maus-Modell generiert oder tatsächlich auch Patientendaten prospektiv erhebt. Also das wär' sicher was, was ich mir gewünscht

hätte, was aber in dem Umfang nicht da war. Also, und explizite Schulungen zum Bereich Forschungsdatenmanagement, das war ja die Frage, ähm.

43. **B:** Wusst ich/ Also, ich seh jetzt schon, seitdem ich dieses Angebot eingeholt habe, seh' ich, bin ich jetzt auch im E-Mail-Newsletter und sehe ab und an, dass es Angebote gibt, ja. Ähm muss aber fairerweise gestehen, dass ich diese noch nicht wahrgenommen hab'. Aber das sind dann eher Vortragsreihen oder sowas. Konkrete Schulungen irgendwie für vielleicht Junior Scientists oder so ähnlich habe ich in der Form jetzt noch nicht gesehen, nein. Beratung sicherlich. Beratung hab' ich bekommen, aber Schulungen eher weniger und der Forschungsdatenmanagementplan oder die Vorlage dessen war, ja, ... ausbaufähig.
44. **I:** OK, kommen wir jetzt eigentlich schon zu meiner letzten Frage. [...] Wenn alles möglich wäre, welche Services oder welche Unterstützung seitens der Universität würdest du dir wünschen bei der Erstellung von Metadaten für deine Forschungsdaten?
45. **B:** Bei der Erstellung von Metadaten für meine Forschungsdaten, ja. Ähm, ja, also das was ich jetzt eigentlich eben schon angesprochen habe. Ein auch in Kooperation mit der IT, dass es Databases gibt. Ähm, dass es sozusagen ja irgendwie Register Vorlagen gibt gespeichert auf den Servern der Universität, die man nutzen kann, um Patientendaten einzutragen. Wo dann automatisch auch Daten generiert werden zu: wann hab' ich das eingetragen, wer hat dann noch Zugriffsrechte und so weiter und so fort. Ähm ... jetzt nicht nur zu den Metadaten, sondern halt auch vielleicht noch ein bisschen mehr ähm Unterstützung bei der Erstellung von Forschungsdatenmanagementplänen für... auch dann Einreichung vielleicht auch denn von größeren internationalen Projekten, wo das ja auch zunehmend wichtig ist, dass man beschreibt, welche Daten man generiert und wie man mit diesen umgeht. Ähm und natürlich denn auch mit den Metadaten umgeht. Ähm, ja, das würd' ich mir wünschen. Einfach Strukturen, die einfach verständlich sind, zugänglich sind, reproduzierbar sind. Am besten Portale, die alle gleich benutzen, dass man sich nicht immer in neue Software einarbeiten muss, sondern dass alle oder viele Register einfach mit der gleichen Databasis arbeiten und da halt die entsprechenden Metadaten generiert werden. Ähm, und die dann auch gut reproduzierbar sind und teilbar sind, wenn man das möchte. Das würd' ich mir wünschen, ja.

46. **I:** Gibt es noch irgendwas, dass du zu diesem Thema, Forschungsdatenmanagement, sagen möchtest, dass ich vielleicht nicht gefragt habe?

47. **B:** Tut mir leid, nein, das wär' alles.

48. [...]