

Metadata in Psychology: What researchers really need

Study description of the data referring to the expert interviews conducted in the BMBF-funded project PsyCuraDat

PsyCuraDat Expert Interviews

Version 1.0.0

Katarina Blask, Maria Jalynskij, and Lea Gerhards

Leibniz-Institute for Psychology Information (ZPID)



Content

Figures and Tables	3
1. Preface	4
2. Study information	4
3. Conceptualization of the study	4
3.1 The PsyCuraDat project	4
3.2 The expert interviews	5
4. Content of the study	5
5. Sampling procedure	6
6. Selection procedure	6
7. Data preparation	7
8. Data analysis	14
9. Links	21
Appendix	22

Figures and Tables

Figure 1. Absolute frequencies of reuse frequency	15
Figure 2. Absolute frequencies of reuse purposes mentioned by the interviewees	16
Figure 3. Absolute frequencies of metadata relevant to an optimized reuse of psychological research data.	17
Figure 4. Absolute frequencies of metadata provided during upload	19
Figure 5. Absolute frequencies of metadata that should be included in a standard for psychological research data according to the interviewees	20
Table 1: Interview duration	7
Table 2: Variables in the dataset	8
Table 3: Metadata	8
Table 4: Paradata	9
Table 5: Substantial data	9
Table 6: Missing values	14
Table 7: Absolute frequencies of best and worst documented data	18

1. Preface

This publication and the corresponding dataset refer to data collected within the project „Development of user-oriented curation criteria for psychological research data“ (PsyCuraDat), which is funded by the Federal Ministry of Education and Research (BMBF). Despite careful examination of the data, mistakes may remain undiscovered. If you find mistakes, please let us know by writing an e-mail to PsyCuraDat@leibniz-psychology.org. We will correct these errors in a timely manner and upload a corrected version of the study.

We recommend researchers to always work with the latest version of the PsyCuraDat datasets. You can find them in the disciplinary repository PsychArchives under the DOI: <http://dx.doi.org/10.23668/psycharchives.2756>

If your scientific publication refers to the data of the PsyCuraDat project, we would be pleased to hear from you (i.e. receive bibliographic details of your publication) via e-mail to PsyCuraDat@leibniz-psychology.org.

Our data are subject to the license [CC-BY-SA 4.0 International](#). If you refer to these data from the PsyCuraDat project in your publications, please use the following reference:

Blask, K., Jalynskij, M., & Gerhards, L. (2020). BMBF-project PsyCuraDat: Expert interviews (Metadata in Psychology 1.0: What researchers really need). Data repository PsychArchives. PsyCuraDat: Expert interviews, interview transcripts and data file version 1.0.0, doi: <http://dx.doi.org/10.23668/psycharchives.2756>.

2. Study information

- Study number: PsyCuraDat Expert Interviews
- Study title: Metadata in Psychology: What researchers really need
- Version: 1.0.0, 07.02.2020, doi: <http://dx.doi.org/10.23668/psycharchives.2756>.
- Survey period: 07.10.2019 – 10.01.2020
- Funder: BMBF, funding code: 16QK08

3. Conceptualization of the study

3.1 The PsyCuraDat project

The Open Science Movement becoming general practice in the empirical sciences and thus particularly in psychology has generally two aims: On the one hand, it aims at improving the research economy and on the other hand the quality of scientific work. As a result, not only the

calls for open access as the free availability of the research outcome, i.e. the publication, but also as the free availability of scientific data are growing louder.

However, the goal of preserving the long-term interpretability of research data requires substantial effort from researchers. This is because psychology is a complex field with multiple assignments to the social sciences, humanities, and natural sciences with correspondingly diverse, hardly standardized qualitative and quantitative methods. Common standards for the documentation of psychological research data that can meet these discipline- and method-specific requirements and thus guarantee the long-term interpretability and reusability of these data are largely lacking at the moment.

Even the PsychData documentation standard developed by ZPID, which is meant to ensure the findability and interpretability of significant psychological studies, is not very well known in the field. This might be seen as an indicator of insufficient usability as well as a mismatch between the requirements of a high-quality documentation and the method-specific requirements grounded in the reuse of psychological research data. To bridge the gap between these different requirements, the project PsyCuraDat aims at the development of user-oriented curation criteria considering the needs of researchers in their role as contributors and users of research data. Thus, the overarching goal is to enable a more effective and efficient documentation and reuse of psychological research data.

3.2 The expert interviews

One central element, if one wants to develop user-oriented curation criteria for psychological research data, is the exploration of researchers' needs. In particular, a documentation standard for psychological research data that is oriented towards discipline-specific methods for the reuse of these data requires, among other things, interviewing experts for these methods. Therefore, ten semi-structured interviews were conducted with experts in different psychological methods (e.g. meta-analysis, simulation studies, re-analyses).

4. Content of the study

Experts were interviewed on twelve topics, eight addressing secondary data use from the perspective of a data user and four from the perspective of a data provider:

I. Secondary data use from the perspective of a data user

1. Frequency of secondary data use compared to primary data use
2. Frequent reuse purposes

3. Metadata needed for optimized reuse
4. Further reuse methods and related metadata
5. Data types of used secondary data
6. Best documented data types
7. Worst documented data types
8. (Assumed) reasons for differences in the documentation quality

II. Secondary data use from the perspective of a data provider

9. Metadata provided for uploaded data
10. Sufficiency of the provided metadata
11. Metadata standards used for documentation
12. Most important metadata for a data documentation standard based on JARS

The answers given to the related questions were coded by the PsyCuraDat project staff in different response categories constituting the variables of the present dataset.

The interview guide can be found in appendix (A) of this study description.

5. Sampling procedure

- Study area: Europe, Asia, and USA
- Basic population and selected population: The population for the expert interviews were psychological researchers from universities and research institutes in different European, Asian and US-American countries. In virtue of the methodological approach of the project, only experts in the reuse of psychological research data constituted the selected sample.

6. Selection procedure

- Sample selection: Arbitrary selection; that is only researchers that agreed to be interviewed in the context of our project were selected. The consent forms for the collection and processing of personal data as well as for the transfer and use of personal data for scientific purposes after the end of the project can be found in appendix (B) of this study description.
- Interviewer selection: All interviews were conducted by the project member Katarina Blask.

- Survey procedure: Paper-and-Pencil-Interview (PAPI) with semi-structured interview guide, audio recording and subsequent transcription of the interviews
- Data collection: Via a project member of the PsyCuraDat project (Katarina Blask)
- Software used for data collection, archiving and analysis: The interview was conducted via Google Hangouts and the audio files were created via OBS studio; the interview guide for the PAPI and the interview transcription was done with the help of Microsoft Word and Easytranscript 2.50.5; the transfer of the transcript, (meta-)data collection as well as the coding of the open-ended questions were done in the virtual research environment FuD (www.fud.uni-trier.de); processing and analysis of the data was done with IBM SPSS Statistics (Version 26).
- Interview duration:

Table 1

Interview duration

	N	Min.	Max.	Arithmetic Mean
Interview duration in minutes	10*	23	52	35.33

Note. *There was a total of 10 interviews conducted in the project. In one case an audio file could not be recorded because of technical problems. Therefore, the actual duration for this interview is unknown. (Code -98 'Error in data' in the data file). In two other cases the recording started only with a minimal delay, so that these interviews are nevertheless included in this list. However, due to these recording errors and the statistical nature of the indicators (i.e. minimum, maximum, and arithmetic mean), it is recommended to interpret these values only as estimators of the actual interview duration.

7. Data preparation

- Data control: Control for completeness of the variables in the dataset, correctness of variable names and labels, value labels as well as missing values
- Units in the dataset: 10 realized interviews
- Variables in the dataset (for an explanation, see below):

Table 2

Variables in the dataset

Variable type	Number
Metadata	5
Paradata	5
Substantial data	60

- Metadata: Serve the description of the dataset; these data comprise the PsyCuraDat study number, version, DOI as well as survey year and period

Table 3

Metadata

Variable	Label
study	study number
version	archived version
doi	Digital Object Identifier
year	survey year
period	survey period

- Paradata: Refer to information resulting from the survey process (e.g. interview duration) and administrative information

Table 4

Paradata

Variable	Label
id	ID interviewee
intdat	interview date
begin	beginning of the interview
end	end of the interview
intdura	interview duration in minutes

- Substantial data: Refer to the topics listed under point 4; the values of all variables represent the frequency with which respondents named the different categories

Table 5

Substantial data

Variable	Label
frequency_1a	only reuse
frequency_1b	creation of primary data outweighs reuse
frequency_1c	creation of primary data and reuse are equal
purposes_2a	answering new research questions
purposes_2b	examination of the reproducibility
purposes_2c	meta-analyses

purposes_2d	illustration of methods
purposes_2e	re-analyses
purposes_2f	teaching
purposes_2g	dimension reduction (e.g. principal component analysis)
purposes_2h	simulation studies
purposes_2i	methodological research
metadata_opti_3a	codebook
metadata_opti_3b	information on study quality/reliability
metadata_opti_3c	study protocol
metadata_opti_3d	scripts
metadata_opti_3e	statistical characteristics
metadata_opti_3f	machine readable/standardized metadata
metadata_opti_3g	links
further_methods_4a	None
further_methods_4b	answering new research questions
further_methods_4c	combining data from studies examining similar research questions
data_types_5a	archival data

data_types_5b	behavioral data
data_types_5c	physiological and psychobiological data
data_types_5d	questionnaire data
data_types_5e	biomedical data
documentation_best_6a	one's own data
documentation_best_6b	recent studies
documentation_best_6c	PISA study
documentation_best_6d	bio-medical/psychobiological data are better than behavioral data
documentation_best_6e	behavioral data
documentation_best_6f	self-reports and survey data
documentation_worst_7a	applied studies
documentation_worst_7b	old studies
documentation_worst_7c	behavioral data are worse than bio-medical/physiological data
documentation_worst_7d	EEG data
documentation_worst_7e	behavioral empirical data
documentation_worst_7f	qualitative studies
reasons_8a	individual reasons

reasons_8b	research infrastructure
reasons_8c	(lack of) discipline-specific RDM culture
reasons_8d	no differences in the documentation quality
reasons_8e	complexity of the data
metadata_upload_9a	responsibilities
metadata_upload_9b	study protocol
metadata_upload_9c	list with data file names and descriptions
metadata_upload_9d	effect sizes
metadata_upload_9e	scripts
metadata_upload_9f	codebook
metadata_upload_9g	provenance
metadata_upload_9h	coding scheme
sufficiency_10a	yes
sufficiency_10b	no
sufficiency_10c	actually yes, but also dependent on data user
sufficiency_10d	estimation not possible
metadata_standards_11a	faculties' own checklists
metadata_standards_11b	no use of any known standards

metadata_standards_11c	Ruben Arslan's codebook
metadata_standards_11d	orientation towards the community
standard_JARS_12a	codebook
standard_JARS_12b	code/syntax, data preparation and analyses scripts
standard_JARS_12c	study protocol
standard_JARS_12d	standardized metadata
standard_JARS_12e	rights (e.g.licensing)

- Coding of open-ended questions: All questions within the interview were open-ended questions. To guarantee the inter-rater reliability, the responses to these questions were independently coded by the three project members. The coding was conducted by using a category scheme, which was based on the interview guide and could flexibly expanded by further categories. The basis for this were the previously written transcripts of the interviews. After all interviews had been coded, categorizations were compared by the three project members. In case of dissenting assignments these were discussed by the project members and set to the consensually determined category. The final category scheme is defined by the substantial variables contained in the dataset. In virtue of data protection requirements, the archived interview transcripts can only be provided in an anonymized form and after signature of a user agreement. If you are interested in the data, please contact PsyCuraDat@leibniz-psychology.org.

- Missing values: Assignment of negative values according to the following scheme:

Table 6

Missing values

Code	Label
-99	not mentioned
-98	error in the data

- Analysis system/format: The dataset is available in the format .sav for the system SPSS and as a .csv file.

8. Data analysis

Given that the measurement level of all substantial variables is nominal, the analysis was limited to a descriptive representation of the absolute frequencies for the different response categories.

Considering this representation for the first category, i.e. frequency of reuse, it turns out that for eight out of the ten interviewees, primary data production outweighs the reuse of data from other researchers or even the own lab (see Figure 1).

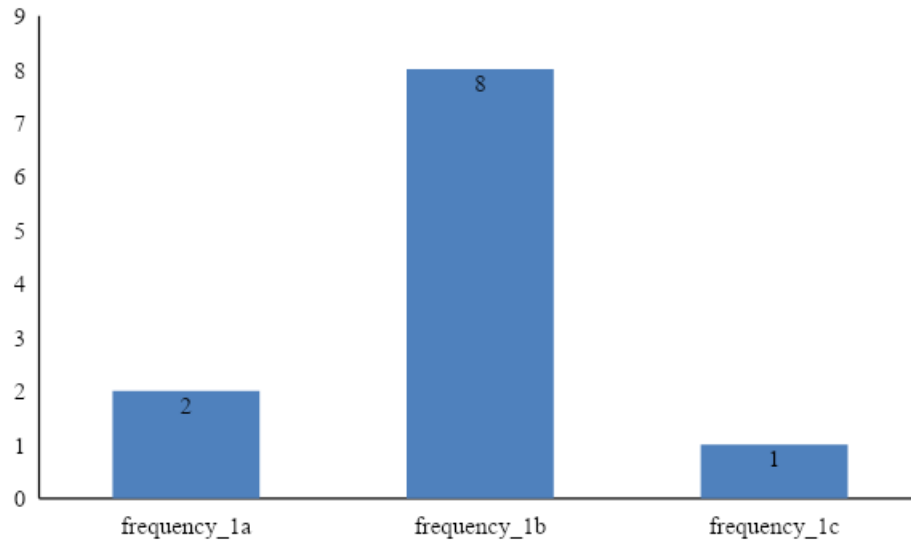


Figure 1. Absolute frequencies of reuse frequency. frequency_1a = only reuse, frequency_1b = creation of primary data outweighs reuse, frequency_1c = creation of primary data and reuse are equal.

Regarding the purposes for which the interviewed researchers generally reuse data, Figure 2 shows that the most common reuse purposes are answering new research questions (*purposes_2a*), meta-analysis (*purposes_2c*), and re-analyses (*purposes_2e*).

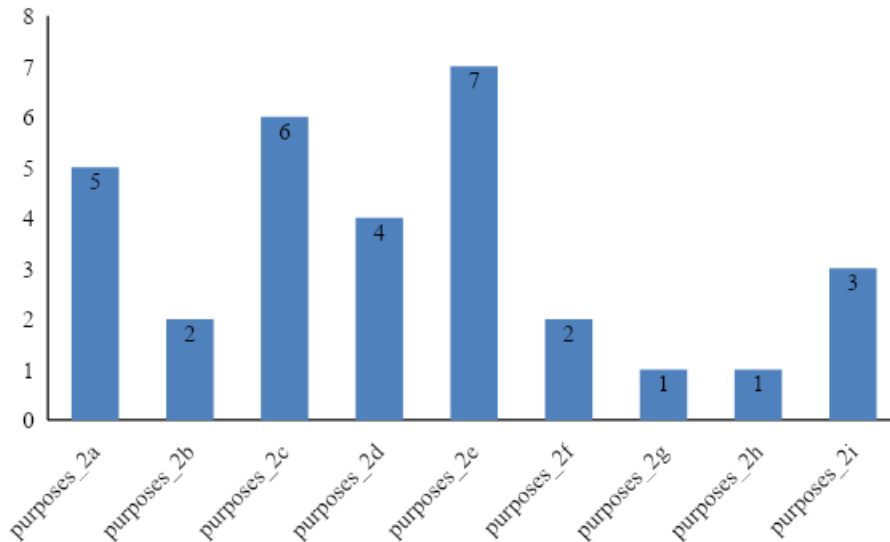


Figure 2. Absolute frequencies of reuse purposes mentioned by the interviewees. purposes_2a = answering new research questions, purposes_2b = examination of the reproducibility, purposes_2c = meta-analysis, purposes_2d = illustration of methods, purposes_2e = re-analyses, purposes_2f = teaching, purposes_2g = dimension reduction, purposes_2h = simulation studies, and purposes_2i = methodological research.

When asked to indicate the metadata that would be most valuable for them to optimize their work with regard to the aforementioned reuse purposes, the majority of interviewees mentioned a codebook, analyses and data preparation scripts, and something like a study protocol (for a complete overview see Figure 3). In this context, the codebook describes the layout of the data in the data file and what the data codes mean. The scripts describe the processing of the data codes, and the study protocol contains all procedural information related to the data generation process (e.g. study goals and objectives, methodology, design, etc.). Preferably the latter should be written before the data collection process starts. Note that only a prospective formulation of a study protocol promotes transparency and thus research integrity in psychology. Additionally, it should be mentioned that these metadata, mentioned in the context of optimized reuse, were also perceived as sufficient for the other reuse purposes which the interviewed researchers know, but have not used by their own yet (i.e. answering new research questions; combining data from studies examining similar research questions).

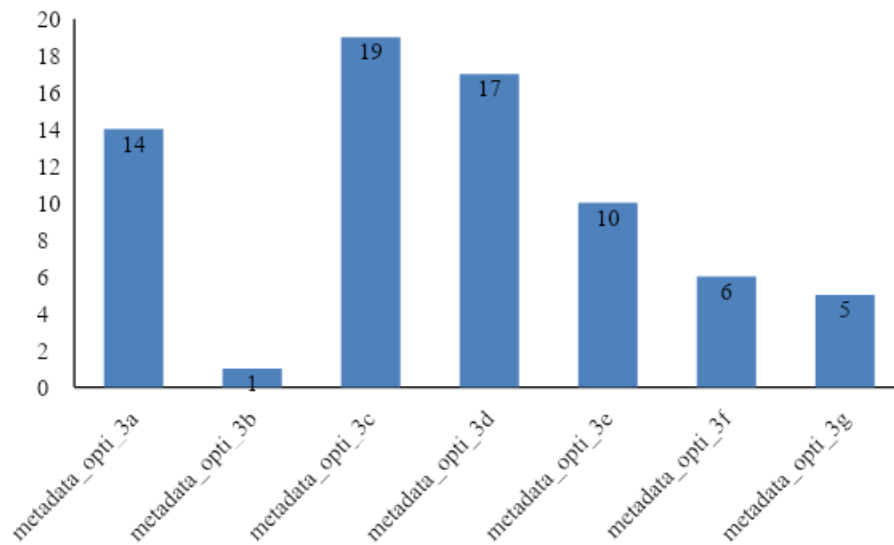


Figure 3. Absolute frequencies of metadata relevant to an optimized reuse of psychological research data. metadata_opti_3a = codebook, metadata_opti_3b = information on study quality/reliability, metadata_opti_3c = study-protocol, metadata_opti_3d = scripts, metadata_opti_3e = statistical characteristics, metadata_opti_3f = machine-readable/standardized metadata, and metadata_opti_3g = links.

Regarding the types of the reused data and their documentation quality, two things have to be stated. First, the interviewed researchers reused behavioral data disproportionately often compared to other data types, such as physiological and psychobiological data. However, the documentation quality of behavioral data is often perceived to be quite low. Thus, here seems to be a discrepancy between researchers' need to reuse behavioral data and the ease of reuse, which is determined by the documentation quality. Second, the documentation quality is strongly determined by the recency of the data. That is, the data documentation of recent studies is perceived to be of higher quality than that of older studies (see Table 7 for an overview on the perceived documentation quality). When the interviewees were asked to surmise the reasons for these discrepancies in documentation quality, the majority attributed them to a lack of a discipline-specific research data management culture ($N = 7$). But also individual reasons ($N = 3$), such as laziness, and an insufficient infrastructure ($N = 4$) were indicated as possible reasons.

Table 7

Absolute frequencies of best and worst documented data

Data types	Documentation best	Documentation worst
one's own data	1	
data from the latest studies	3	
self-reports and survey data	4	
biomedical/psychobiological data	2	1
behavioral data		6
data from old studies		3
data from applied studies		1
data from qualitative studies		1

In sum, researchers seem to have a quite clear idea of how data should be documented for an optimal reuse. However, these ideas are not reflected in present data documentation practices because a clear consensus on how to document psychological research data is missing. The lack of a viable standard for research data documentation also becomes apparent in the diversity of the metadata generally provided during upload and their perceived sufficiency. Indeed, most of the interviewed researchers provide scripts when they upload their data. However, only half of the researchers supply information that might be summarized in a study protocol or a comprehensive codebook (see Figure 4 for a complete overview). Thus, there seems to be a clear attitude-behavior gap. Interestingly, researchers are aware of this gap, but do not act accordingly. Actually, half of the interviewed researchers think that their data documentation does not allow an optimal reuse of their data.

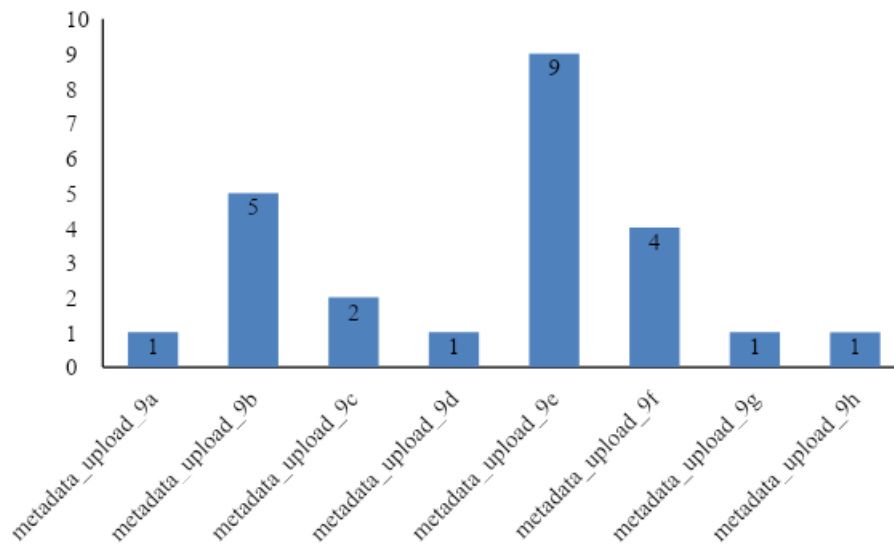


Figure 4. Absolute frequencies of metadata provided during upload of research data. metadata_upload_9a = responsibilities, metadata_upload_9b = study protocol, metadata_upload_9c = list with data file names and descriptions, metadata_upload_9d = effect sizes, metadata_upload_9e = scripts, metadata_upload_9f = codebook, metadata_upload_9g = provenance, and metadata_upload_9h = coding scheme.

This attitude-behavior gap is probably grounded in the absence of a discipline-specific data documentation standard for all psychological research data. That the existing standards are obviously neither sufficient nor appropriate for the documentation of psychological research data is indicated by the answers given in response to the last two questions of the interview. In particular, eight of the ten interviewed researchers have never used an existing standard for documenting their data, indicating that present standards are not perceived as part of the discipline-specific data documentation culture. Finally, researchers' idea of a documentation standard (i.e. study protocol, codebook, and scripts) does not match any of the existing standards. For an overview of the information that should be included in a standard for psychological research data according to the interviewees, see Figure 5.

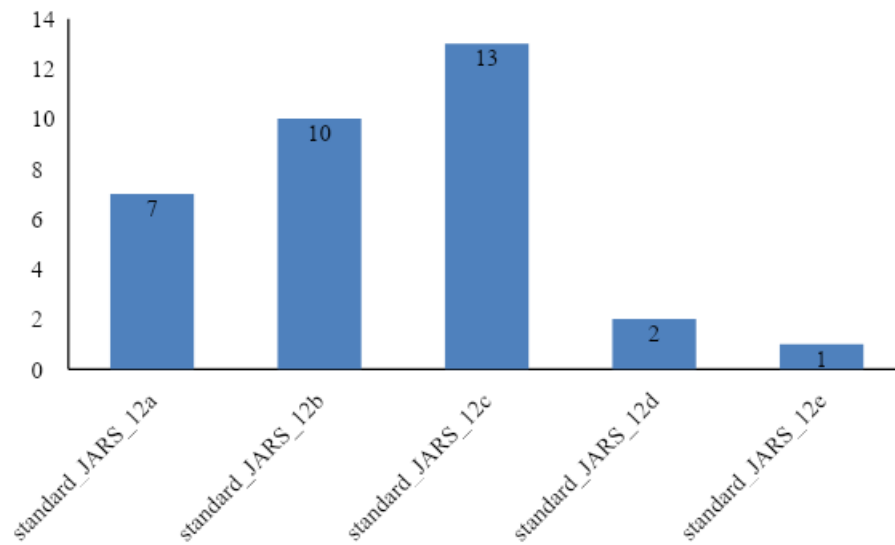


Figure 5. Absolute frequencies of metadata that should be included in a standard for psychological research data according to the interviewees. standard_JARS_12a = codebook, standard_JARS_12b = code/syntax, data preparation and analyses scripts, standard_JARS_12c = study protocol, standard_JARS_12d = standardized metadata, and standard_JARS_12e = rights (e.g. licensing).

9. Links

PsyCuraDat project website: www.leibniz-psychology.org/institut/drittmittelprojekte/psycuradat

Leibniz Institute for Psychology Information (ZPID): www.leibniz-psychology.org

Virtual Research Environment FuD: www.fud.uni-trier.de

Appendix

A. Interview guide “Expert interviews PsyCuraDat” (German & English version)

German version:

1. Begrüßung

Hallo Herr XXX! Zunächst möchte ich Ihnen nochmals dafür danken, dass Sie uns in unserem Projekt unterstützen und an diesem Interview teilnehmen. Bevor wir beginnen, würde ich Ihnen gerne noch einmal unsere Ziele und die damit verbundene Forschungsfrage erläutern.

2. Erklären der Ziele, Fragestellung und Vorgehensweise

Übergeordnetes Ziel unseres Forschungsprojekts ist die Entwicklung eines Dokumentationsstandards, welcher die Nachnutzbarkeit psychologischer Forschungsdaten optimieren soll. Ausgehend von diesem übergeordneten Ziel, sind wir daran interessiert zu erfahren, was Forschende brauchen, wenn sie psychologische Forschungsdaten nachnutzen.

3. Beschreibung des Status Quo bezüglich der Nachnutzung von Forschungsdaten (aus Perspektive eines Datennutzers)

Zunächst möchten wir die Nutzung von Sekundärdaten aus Ihrer Sicht als Datennutzer etwas näher beleuchten.

1. Wie oft haben Sie in der Vergangenheit Datensätze aus Ihrem Labor und anderen Labors wiederverwendet? Könnten Sie das Ganze quantifizieren, indem Sie uns die relative Häufigkeit nennen mit der Sie Daten wiederverwenden im Vergleich zur Erzeugung von eigenen Primärdaten?
2. Für welche Zwecke haben Sie in der Vergangenheit Sekundärdaten verwendet?
 - a. Welche spezifischen Zusatzinformationen zu diesen Daten (z.B. Metadaten) benötigen Sie, um Ihre Arbeit mit diesen Sekundärdaten zu optimieren?
 - b. Gibt es andere Methoden der Daten-Nachnutzung (die Sie kennen, die Sie aber nicht selbst verwendet haben), die andere Metadaten erfordern würden als die, die Sie bereits erwähnt haben?

3. Welche Art von Daten verwenden Sie im Allgemeinen für die verschiedenen Zwecke (z.B. physiologische oder verhaltensbezogene Daten, Video- oder andere Materialien, die in der psychologischen Forschung verwendet werden)?
 - a. Welche dieser Daten sind aus Ihrer Sicht die am besten dokumentierten und welche die am schlechtesten dokumentierten?
 - b. Wo liegen Ihrer Meinung nach die Gründe für diese Unterschiede in der Dokumentationsqualität?
4. Beschreibung des Status Quo bezüglich der Nachnutzung von Forschungsdaten (aus Perspektive eines Datengebers)

Nun aus der Sicht eines Datengebers:

1. Welche Art von Metadaten stellen Sie allgemein über einen Datensatz bereit, wenn Sie ihn hochladen?
2. Sind Sie der Meinung, dass diese Metadaten ausreichen, um Ihren Datensatz nachzunutzen?
3. Haben Sie bestimmte Metadatenstandards zur Dokumentation Ihrer Daten verwendet?
 - a. Wenn ja, welche Metadaten fehlen in diesen Standards, die für eine verbesserte Nachnutzbarkeit psychologischer Forschungsdaten einbezogen werden sollten (z.B. Informationen über die experimentellen Bedingungen, Hauptanalysen, Zusatzanalysen)?
4. Wenn Sie einen Metadatenstandard erstellen würden, was sind Ihrer Meinung nach die wichtigsten Informationen? Welche zusätzlichen Informationen zu Ihren Daten sind für eine optimale Nachnutzung am wichtigsten? Vielleicht können Sie über diese Frage mit Bezug auf die Abschnitte nachdenken, die in den Journal Article Reporting Standards der APA enthalten sind, d.h. alle Informationen, die im Allgemeinen in einem Forschungsartikel enthalten sind.

Ende des Interviews

Haben Sie weitere Fragen, Bedenken oder Kommentare, die Ihrer Meinung nach nützlich sein könnten oder die Sie hervorheben möchten?

Wenn Sie keine Fragen mehr haben, möchte ich mich noch einmal für Ihre Teilnahme am Interview bedanken. Sie helfen uns mit Ihrer Teilnahme dabei, einen Metadatenstandard zu entwickeln, der hoffentlich für die psychologische Fachgemeinschaft nützlich sein wird! Sollten Sie weitere Kommentare oder Ideen haben, können Sie mir gerne eine E-Mail schreiben.

English version:

1. Welcoming

Hello Mr/Mrs XXX! First of all, I would like to thank you once more for supporting our project by taking part in this interview. Before we start, I would like to briefly introduce our goals and the related research questions.

2. Explaining the goals and the research question

The overarching goal of our research project is the development of a documentation standard for psychological research data to optimize their reusability. With this focus on reuse in mind, we are starting from the viewpoint of what people want and need when they (try to) reuse psychological data.

3. Description of the status quo regarding secondary data use (from the perspective of a data user)

First of all, we would like to elaborate on secondary data use from your perspective as a data user.

1. How often you reused datasets from your lab and other labs in the past? Could you quantify your specification by providing the relative frequency for reusing data compared to producing primary data?
2. For which purposes have you used secondary data in the past?
 - a. What specific additional information on that data (i.e., metadata) would you need to optimize that work?
 - b. Are there other methods of secondary data use (you know about, but have not used on your own) that would require other metadata than the ones you already mentioned?
3. What kind of data are you generally using for the different purposes (e.g., physiological, behavioral, video, or other materials used in psychological research)?

- a. From your perspective, which of these data are best and which worst documented?
 - b. Where do you think are the reasons for these differences in documentation quality?
4. Description of the status quo regarding secondary data use (from the perspective of a data provider)

Now taking the perspective of a data provider:

1. What sorts of metadata do you generally provide about a dataset when you upload?
2. Do you think that these metadata are sufficient for reusing your dataset?
3. Have you used certain metadata standards for annotating your data?
 - a. If yes, what metadata is missing from these standards that should be incorporated for an improved reusability of psychological research data (e.g., information about the experimental conditions, main analyses, additional analyses)?
4. If you were to create a metadata standard, what do you think is the most important information? That is which additional information on your data is most important for an optimal reuse? Perhaps you can think of this question in terms of the sections included in the Journal Article Reporting Standards from the APA, i.e., all the information generally included in a research article.

Ending the Interview

Do you have any further questions, concerns or comments that you think can be useful or you want to emphasize?

Ok then, if you have no more questions, once again I would like to thank you for taking part in the interview and helping us develop what we hope will be a useful standard for the psychological community! If you have any further comments or ideas, feel free to email me.

B. Consent forms PsyCuraDat “Expert interviews” (German & English version)

German version of the consent forms:

1. Einwilligungserklärung zur Erhebung und Verarbeitung personenbezogener Interviewdaten (nach Vorlage der Arbeitsgruppe „Datenschutz und Qualitative Sozialforschung“ (RatSWD))

Forschungsprojekt: PsyCuraDat

Durchführende Institution: Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID)

Projektleitung: PD Dr. Erich Weichselgartner

Interviewerin/Interviewer: Dr. Katarina Blask

Interviewdatum:

Beschreibung des Forschungsprojektes (Zutreffendes bitte ankreuzen):

☐ mündliche Erläuterung

☒ schriftliche Erläuterung

Projektbeschreibung

Die in den empirischen Wissenschaften und somit auch speziell in der Psychologie um sich greifende Open-Science-Bewegung verfolgt grundsätzlich zwei Ziele. Zum einen soll die Forschungsökonomie verbessert werden und zum anderen die Qualität wissenschaftlichen Arbeitens. Im Zuge dessen wurde nicht nur der Ruf nach Open Access als die freie Verfügbarkeit des Forschungs-Endprodukts /Publikation/, sondern auch die Forderung nach der freien Verfügbarkeit wissenschaftlicher Daten zunehmend lauter.

Das Ziel, die langfristige Interpretierbarkeit der Forschungsdaten zu erhalten, bedeutet jedoch einen erheblichen Arbeitsaufwand, denn die Psychologie ist ein komplexes Fach mit multipler Zuordnung zu den Sozial-, Geistes- und Naturwissenschaften und entsprechend diversen, wenig standardisierten, qualitativen und quantitativen Methoden. Einheitliche Standards zur Dokumentation psychologischer Forschungsdaten, welche diesen fach- und methodenspezifischen Anforderungen entsprechen können und damit die langfristige Interpretierbarkeit und Nutzbarkeit dieser Daten gewährleisten, fehlen bislang jedoch weitestgehend.

Auch der vom ZPID entwickelte PsychData-Dokumentationsstandard, welcher die Auffindbarkeit und Interpretierbarkeit bedeutsamer psychologischer Studien gewährleisten sollte, fand keine weite Verbreitung im Fach. Dies kann sowohl als Indiz für eine mangelnde Nutzerfreundlichkeit gesehen werden als auch für eine unzureichende Passung zwischen den Anforderungen an eine qualitativ hochwertige Dokumentation und den methodenspezifischen Anforderungen einer Nachnutzung psychologischer Forschungsdaten. In dem Projekt PsyCuraDat wird deshalb die Entwicklung nutzerorientierter Kurationskriterien angestrebt, welche die Bedürfnisse der Forschenden in ihrer Rolle als Datengeber und Datennehmer berücksichtigen. Darüber soll eine effizientere und effektivere Dokumentation und Nachnutzung psychologischer Forschungsdaten ermöglicht werden.

Um dieses Projektziel erreichen zu können, werden in einem ersten Schritt Experteninterviews mit Forschenden aus der Psychologie durchgeführt. Die Forschenden zeichnen sich durch ihre besonderen Kenntnisse in spezifischen Methoden aus, welche auf die Nachnutzung bestehender Forschungsdaten abzielen (z.B. Meta-Analysen und Replikationen).

Durchführung des Interviews

Die Interviews werden mittels GoogleHangouts durchgeführt und über die Software OBS Studio auditiv aufgezeichnet und sodann von den Mitarbeiterinnen und Mitarbeitern des Forschungsprojektes PsyCuraDat in Schriftform gebracht. Für die weitere wissenschaftliche Auswertung der Interviewtexte werden alle Angaben, die zu einer Identifizierung der Person führen könnten, verändert oder aus dem Text entfernt. In wissenschaftlichen Veröffentlichungen werden Interviews nur in Ausschnitten zitiert, um gegenüber Dritten sicherzustellen, dass der entstehende Gesamtzusammenhang von Ereignissen nicht zu einer Identifizierung der Person führen kann.

Personenbezogene Kontaktdaten werden von Interviewdaten getrennt für Dritte unzugänglich gespeichert. Nach Beendigung des Forschungsprojektes werden Ihre Kontaktdaten automatisch gelöscht, es sei denn, Sie stimmen einer weiteren Speicherung zur Kontaktmöglichkeit für themenverwandte Forschungsprojekte ausdrücklich zu. Selbstverständlich können Sie einer längeren Speicherung zu jedem Zeitpunkt widersprechen.

Die Teilnahme an dem Interview ist freiwillig. Sie haben zu jeder Zeit die Möglichkeit, das Interview abubrechen und Ihr Einverständnis in eine Aufzeichnung und Niederschrift der Interviews zurückzuziehen, ohne dass Ihnen dadurch irgendwelche Nachteile entstehen.

Ich bin damit einverstanden, im Rahmen des genannten Forschungsprojektes an einem Interview teilzunehmen.

☐ ja ☐ nein

Ich bin damit einverstanden, für zukünftige themenverwandte Forschungsprojekte kontaktiert zu werden. Hierzu bleiben meine Kontaktdaten über das Ende des Forschungsprojektes hinaus gespeichert.

☐ ja ☐ nein

Vorname, Nachname in Druckschrift

Ort, Datum / Unterschrift

2. Einwilligungserklärung zur Übermittlung und Nutzung personenbezogener Daten für wissenschaftliche Zwecke nach Projektende (nach Vorlage der Arbeitsgruppe „Datenschutz und Qualitative Sozialforschung“ (RatSWD))

Ihre Interviewdaten werden in anonymisierter Form zur Archivierung und weiteren wissenschaftlichen Nutzung an das Forschungsdatenrepositorium PsychArchives übermittelt. Die Vollständigkeit der Anonymisierung der Interviews wird von den Mitarbeitenden des PsyCuraDat-Projektes geprüft. Der Originaltext des Interviews wird nach den genannten Maßnahmen zur Anonymisierung der Interviews gelöscht.

Das Forschungsdatenrepositorium PsychArchives stellt Wissenschaftlerinnen und Wissenschaftlern die anonymisierten Interviewdaten für ausschließlich wissenschaftliche Zwecke zur Verfügung.

Ich bin damit einverstanden.

☐ ja ☐ nein

Um anderen Wissenschaftlerinnen und Wissenschaftlern die Möglichkeit zu geben, Sie für die Teilnahme an weiteren Forschungsvorhaben zur Untersuchung späterer Entwicklungen des Forschungsthemas zu kontaktieren, werden Ihre Kontaktdaten ebenfalls an das Forschungsdatenrepositorium PsychArchives übermittelt. Ihre Kontaktdaten werden von den Interviewdaten getrennt und für Dritte unzugänglich gespeichert. Selbstverständlich können Sie einer Speicherung Ihrer Kontaktdaten in PsychArchives zu jedem Zeitpunkt widersprechen. Die Übermittlung Ihrer Kontaktdaten an andere Wissenschaftlerinnen und Wissenschaftler erfolgt nur zur Kontaktaufnahme für Interviews zu nicht-kommerziellen, wissenschaftlichen

Zwecken in thematisch verwandten Forschungsbereichen.

Ich bin damit einverstanden.

☐ ja ☐ nein

Vorname, Nachname in Druckschrift

Ort, Datum / Unterschrift

English version of the consent forms:

1. Consent form for the collection and processing of personal data (According to the template of the working group „Datenschutz und Qualitative Sozialforschung [data protection and qualitative social research]“, (RatSWD))

Research Project: PsyCuraDat

Implementing Institute: Leibniz Institute for Psychology Information (ZPID)

Project Manager: PD Dr. Erich Weichselgartner

Interviewer: Dr. Katarina Blask

Interview Date:

Project Description (Mark with a cross where applicable):

☐ Oral Explanation

☒ Written Explanation

Project Description

The Open-Science-Movement becoming general practice in the empirical sciences and thus particularly in Psychology has generally two aims: On the one hand, it is targeted at improving the research economy and on the other hand the quality of scientific work. As a result, not only the calls for open access as the free availability of the research outcome, i.e. the publication, but also as the free availability of scientific data are growing louder.

However, the goal of preserving the long-term interpretability of research data require substantial effort from researchers. This is because psychology is a complex field with multiple assignments to the social sciences, humanities, and natural sciences with correspondingly diverse, hardly standardized qualitative and quantitative methods. Common standards for the documentation of psychological research data that can meet these discipline- and method-specific requirements and thus guarantee the long-term interpretability and re-usability of these data are largely lacking at the moment.

Even the PsychData-documentation standard that was developed by the ZPID to ensure the findability and interpretability of significant psychological studies, is not very well known in the field. This might be seen as an indicator for an insufficient usability as well as a mismatch between the requirements of a high-quality documentation and the method-specific requirements grounded in the re-use of psychological research data. To this end, the project PsyCuraDat aims at the development of user-oriented curation criteria considering the needs of researchers in their role as contributors and users of research data. Thus, the overarching goal is to enable a more effective and efficient documentation and re-use of psychological research data.

To this end, in a first step expert interviews with researchers from psychology will be conducted. All these researchers are characterized by their expert knowledge of specific methods directed towards the re-use of existing research data (e.g., meta-analysis and replications).

Conduct of the interview

The interviews will be conducted via GoogleHangouts and auditively recorded with the software OBS Studio. Afterwards the research associates of the PsyCuraDat Project will transcribe the interviews. For further scientific analysis of the interview transcripts all data that might contribute to an identification of the person will be changed or removed from the text. In scientific publications interviews will only be quoted in sections, to ensure that the emerging general context of events cannot lead to an identification of the person via third parties.

Personal contact details will be saved separately from the interview data and inaccessible for third parties. After finishing the research project your contact details will automatically be deleted, unless you expressly agreed to the further storage of your contact details for subject-related research projects. Of course, you can contradict an extended storage of your data at any time.

The participation in the interview is voluntary. You have the possibility to terminate the interview and to withdraw your consent for the auditory and textual record of the interview at any time without disadvantages.

I agree to participate in an expert interview within the context of the mentioned research project.

☐ yes ☐ no

I agree, to be contacted for future subject-related research projects. Therefore, my contact details will be stored beyond the end of the research project.

☐ yes ☐ no

First name, Last name in print letters

Place, Date / Signature

2. Consent form for the transfer and use of personal data for scientific purposes after the end of the project (According to the template of the working group „Datenschutz und Qualitative Sozialforschung [data protection and qualitative social research]“, (RatSWD))

For archiving and further scientific use your data will be transferred to the research data repository PsychArchives in anonymized form. The completeness of anonymization will be checked by the research associates of the PsyCuraDat project. After anonymization the original interview files will be deleted.

The research data repository PsychArchives provides the anonymized research data to other researchers only for scientific purposes.

I agree.

☐ yes ☐ no

In order to give other researchers the possibility to contact you for participation in further research projects aimed at the investigation of subsequent developments of the research project your contact details will also be transferred to the research data repository PsychArchives. Your contact details will be saved separately from the interview data and inaccessible for third parties. Of course, you can contradict the storage of your contact details in PsychArchives at any time. Your contact details will only be passed to other researchers who want to contact you for interviews with non-commercial, scientific purposes in thematically related research areas.

I agree.

☐ yes ☐ no

First name, Last name in print letters

Place, Date / Signature