

# **Metadata in Psychology 2.0: What researchers really need**

## **Study description of the data referring to the online survey conducted in the BMBF-funded project PsyCuraDat**

PsyCuraDat Online Survey

Version 1.0.0

Katarina Blask, Lea Gerhards, and Maria Jalynskij

Leibniz-Institute for Psychology Information (ZPID)



## Content

Tables.....	3
1. Preface.....	4
2. Study information.....	4
3. Conceptualization of the study .....	5
3.1 The PsyCuraDat project.....	5
3.2 The online survey .....	6
4. Content of the study .....	6
5. Sampling procedure.....	7
6. Selection procedure .....	7
7. Data preparation .....	8
8. Data analysis .....	22
9. References .....	37
10. Links.....	38
Appendix .....	39

## Tables

Table 1: Online survey duration.....	8
Table 2: Variables in the dataset .....	8
Table 3: Metadata.....	9
Table 4: Paradata.....	9
Table 5: Substantial data .....	10
Table 6: Missing values.....	18
Table 7: Origin .....	20
Table 8: Sub-discipline.....	230
Table 9: Means and standard deviations for primary data production and secondary data use .....	21
Table 10: Frequent reuse purposes in Psychology .....	22
Table 11: Metadata needed for different reuse purposes .....	24
Table 12: Contrasts for reuse purposes dependent on the corresponding metadata.....	25
Table 13: Metadata usually provided during upload and their perceived usefulness .....	3430

## 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 inform us by writing an e-mail to [PsyCuraDat@leibniz-psychology.org](mailto:PsyCuraDat@leibniz-psychology.org). We will correct these errors in a timely manner and upload a corrected version of the data.

We recommend researchers always to 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.2751>. If your scientific publications refer 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](mailto:PsyCuraDat@leibniz-psychology.org).

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

Blask, K., Gerhards, L., & Jalynskij, M. (2020): BMBF-project PsyCuraDat: Online Survey (Metadata in Psychology 2.0: What researchers really need). Data repository PsychArchives. PsyCuraDat: Online survey data file version 1.0.0, doi: <http://dx.doi.org/10.23668/psycharchives.2751>.

## 2. Study information

- Study number: PsyCuraDat Online Survey
- Study title: Metadata in Psychology 2.0: What researchers really need
- Version: 1.0.0, 07.02.2020, doi: <http://dx.doi.org/10.23668/psycharchives.2751>.
- Survey period: 06.11.2019 – 12.01.2020
- Funder: BMBF, funding code: 16QK08

### 3. Conceptualization of the study

#### 3.1 The PsyCuraDat project

In the context of the omnipresent Open Science Movement, all stakeholders in psychological research (i.e. researchers, journals, and funders) increasingly strive for more transparency in order to foster scientific exchange and the quality of scientific work. However, to reach this overarching goal of the Open Science Movement, there is a need for discipline-specific documentation standards not only for research articles, but also for the data representing the basis of those articles. Therefore, the project PsyCuraDat has two strategic objectives:

1. Development of a documentation standard for psychological research data considering the discipline-specific methods relevant for reusing these data. This objective is therefore aimed at addressing the needs of researchers in their role as users of research data.
2. Development of a documentation standard for psychological research data considering the needs of psychological researchers in their role as contributors of research data.

In order to specify a documentation standard adhering to these goals on a contextual as well as on a formal level, the project PsyCuraDat pursues three methodological approaches:

1. Conducting interviews with experts of psychological reuse methods (see Blask, Jalynskij, & Gerhards, 2020).
2. Conducting an online survey with psychological researchers from different European countries, the USA, and Canada to broaden the scope of the contextual specification.
3. Conducting a user study to deduce a user-friendly formal specification of the documentation standard and to validate the usefulness of the contextual specification for the relevant reuse methods.

Thus, the overarching goal is to promote research economy and research integrity in Psychology by enabling a more effective and efficient documentation and reuse of psychological research data.

### 3.2 The online survey

The contextual specification of a user-oriented documentation standard for psychological research data is critically dependent on representing researchers' needs in the most objective way possible. To get an impression of researchers' needs in their role as contributors and users of research data, an online survey addressing these topics was conducted. Accordingly, the online survey was aimed at two objectives. The first goal was to test for the reliability of the conclusions derived from the expert interviews, which addressed the same topics (cf. Blask, Jalynskij, & Gerhards, 2020). The second and final goal was to get a more precise picture of researchers' documentation needs by explicitly asking them about the perceived usefulness of metadata that are representative of psychological research methods.

## 4. Content of the study

Comparable to the expert interviews, researchers were asked questions on twelve topics in the survey, 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 in accordance with the response categories defined in the codebook of the EFS questback survey. The response categories belonging to the different questions constitute the substantial variables of the present dataset.

A screenshot of the survey realized in EFS questback can be found in appendix (A) of this study description.

## 5. Sampling procedure

- Study area: Europe, USA, Canada
- Basic population and selected population: The population for the online survey were psychological researchers from universities and research institutes in Europe, the USA, and Canada.

## 6. Selection procedure

- Sample selection: Arbitrary selection; that is only researchers responding to our request sent via an arbitrarily created mailing list (in each case 800 individual researchers or research organizations from Europe, the USA, and Canada, respectively) or whose attention was drawn to our survey by the request shared via the Twitter account of ZPID (Leibniz Institute for Psychology Information) were selected. The request sent to the researchers can be found in appendix (B).
- Survey procedure: Online survey
- Data collection: Via EFS questback
- Software used for data collection, archiving and analysis: The online survey was conducted via EFS questback (Fall, 2019); processing and analysis of the data was done with IBM SPSS Statistics (Version 26).
- Online survey duration:

Table 1

*Online survey duration*

	<i>N</i>	Min	Max	<i>M</i>	<i>SD</i>	<i>M*</i>
Online survey duration in minutes	54	2.33	8863.52	280.78	1383.63	9.40

*Note.* To get a realistic estimation for the time that researchers needed to complete the survey participants deviating more than two standard deviations from the mean were excluded. After excluding these three participants, the mean processing time was 9.40 minutes ( $M^*$  = corrected mean).  $N$  = sample size, Min = minimum, Max = maximum,  $M$  = mean,  $SD$  = standard deviation.

## 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: 63 finished online surveys (including six test trials)
- Variables in the dataset (for an explanation, see below):

Table 2

*Variables in the dataset*

Variable type	Number
Metadata	5
Paradata	16
Substantial data	187

- 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. date and time of survey start) and administrative information

Table 4

*Paradata*

Variable	Label
id	ID participant
persons_ex	excluded participants
datetime	date and time of survey start
date_of_last_access	date of last access
duration	processing duration in seconds
rts3849418	relative timestamp for page 3849418
rts3849714	relative timestamp for page 3849714
rts3850652	relative timestamp for page 3850652

rts3850708	relative timestamp for page 3850708
rts3856413	relative timestamp for page 3856413
rts3858035	relative timestamp for page 3858035
rts3859413	relative timestamp for page 3859413
rts3864259	relative timestamp for page 3864259
rts3868210	relative timestamp for page 3868210
rts3893546	relative timestamp for page 3893546
rts3893556	relative timestamp for page 3893556

- 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
Sex	sex
Origin	origin
discipline	discipline
DisciplineOthers	other than the prescribed disciplines
MainResearchArea	main research area
SecondaryDataUse	secondary data use
PrimaryDataUse	primary data use

MetaAnalysisYourL	meta-analysis from your own lab/organization
MetaAnalysisOtherL	meta-analysis from other labs/organizations
SystematicReviewYourL	systematic review from your own lab/organization
SystematicReviewOtherL	systematic review from other labs/organizations
ReanalysisYourL	reanalysis from your own lab/organization
ReanalysisOtherL	reanalysis from other labs/organizations
IllustrationsYourL	illustrations from your own lab/organization
IllustrationsOtherL	illustrations from other labs/organizations
OthersYourL	other methods from your own lab/organization
OthersOtherL	other methods from other labs/organizations
NoneYourL	no reuse of data from your own lab/organization
NoneOtherL	no reuse of data from other labs/organizations
Method1	other method to reuse data 1
Method2	other method to reuse data 2
Method3	other method to reuse data 3
BehavioralData	behavioral data
PhysiologicalData	physiological data
OthersData	other data types
HypothesesMeta	hypotheses for meta-analyses

HypothesesSystematic	hypotheses for systematic reviews
HypothesesRean	hypotheses for reanalyses
HypothesesIllust	hypotheses for illustrations
HypothesesOthers1	hypotheses for other reuse methods 1
HypothesesOthers2	hypotheses for other reuse methods 2
HypothesesOthers3	hypotheses for other reuse methods 3
SampleSizeMeta	sample size for meta-analyses
SampleSizeSystematic	sample size for systematic reviews
SampleSizeRean	sample size for reanalyses
SampleSizeIllust	sample size for illustrations
SampleSizeOthers1	sample size for other reuse methods 1
SampleSizeOthers2	sample size for other reuse methods 2
SampleSizeOthers3	sample size for other reuse methods 3
CriteriaMeta	inclusion/exclusion criteria for meta-analyses
CriteriaSystematic	inclusion/exclusion criteria for systematic reviews
CriteriaRean	inclusion/exclusion criteria for reanalyses
CriteriaIllust	inclusion/exclusion criteria for illustrations
CriteriaOthers1	inclusion/exclusion criteria for other reuse methods 1
CriteriaOthers2	inclusion/exclusion criteria for other reuse methods 2
CriteriaOthers3	inclusion/exclusion criteria for other reuse methods 3
SettingMeta	setting for meta-analyses
SettingSystematic	setting for systematic reviews

SettingRean	setting for reanalyses
SettingIllust	setting for illustrations
SettingOthers1	setting for other reuse methods 1
SettingOthers2	setting for other reuse methods 2
SettingOthers3	setting for other reuse methods 3
SamplingMMeta	sampling method for meta-analyses
SamplingMSystematic	sampling method for systematic reviews
SamplingMRean	sampling method for reanalyses
SamplingMillust	sampling method for illustrations
SamplingMOthers1	sampling method for other reuse methods 1
SamplingMOthers2	sampling method for other reuse methods 2
SamplingMOthers3	sampling method for other reuse methods 3
ResearchDesignMeta	kind of research design (e.g. correlational, experimental) for meta-analyses
ResearchDesignSystematic	kind of research design (e.g. correlational, experimental) for systematic reviews
ResearchDesignRean	kind of research design (e.g. correlational, experimental) for reanalyses
ResearchDesignIllust	kind of research design (e.g. correlational, experimental) for illustrations
ResearchDesignOthers1	kind of research design (e.g. correlational, experimental) for other reuse methods 1
ResearchDesignOthers2	kind of research design (e.g. correlational, experimental) for other reuse methods 2
ResearchDesignOthers3	kind of research design (e.g. correlational, experimental) for other reuse methods 3

VariablesMeta	included variables (e.g. dependent/independent variables) for meta-analyses
VariablesSystematic	included variables (e.g. dependent/independent variables) for systematic reviews
VariablesRean	included variables (e.g. dependent/independent variables) for reanalyses
VariablesIllust	included variables (e.g. dependent/independent variables) for illustrations
VariablesOthers1	included variables (e.g. dependent/independent variables) for other reuse methods 1
VariablesOthers2	included variables (e.g. dependent/independent variables) for other reuse methods 2
VariablesOthers3	included variables (e.g. dependent/independent variables) for other reuse methods 3
AssignmentMeta	random/nonrandom assignment for meta-analyses
AssignmentSystematic	random/nonrandom assignment for systematic reviews
AssignmentRean	random/nonrandom assignment for reanalyses
AssignmentIllust	random/nonrandom assignment for illustrations
AssignmentOthers1	random/nonrandom assignment for other reuse methods 1

AssignmentOthers2	random/nonrandom assignment for other reuse methods 2
AssignmentOthers3	random/nonrandom assignment for other reuse methods 3
PrimaryAnalysesMeta	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for meta-analyses
PrimaryAnalysesSystematic	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for systematic reviews
PrimaryAnalysesRean	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for reanalyses
PrimaryAnalysesIllust	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for illustrations
PrimaryAnalysesOthers1	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 1
PrimaryAnalysesOthers2	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 2
PrimaryAnalysesOthers3	primary outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 3
AdditionalAnalysesMeta	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for meta-analyses

AdditionalAnalysesSystematic	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for systematic reviews
AdditionalAnalysesRean	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for reanalyses
AdditionalAnalysesIllust	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for illustrations
AdditionalAnalysesOthers1	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 1
AdditionalAnalysesOthers2	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 2
AdditionalAnalysesOthers3	additional outcome analyses (name, classic/Bayesian, effect size, confidence interval) for other reuse methods 3
MediationAnalysesMeta	mediation analyses (mediation approach, significance test) for meta-analyses
MediationAnalysesSystematic	mediation analyses (mediation approach, significance test) for systematic reviews
MediationAnalysesRean	mediation analyses (mediation approach, significance test) for reanalyses
MediationAnalysesIllust	mediation analyses (mediation approach, significance test) for illustrations
MediationAnalysesOthers1	mediation analyses (mediation approach, significance test) for other reuse methods 1



MediationAnalysesOthers2	mediation analyses (mediation approach, significance test) for other reuse methods 2
MediationAnalysesOthers3	mediation analyses (mediation approach, significance test) for other reuse methods 3
MaterialMeta	additional material (information on additional material related to the dataset) for meta-analyses
MaterialSystematic	additional material (information on additional material related to the dataset) for systematic reviews
MaterialRean	additional material (information on additional material related to the dataset) for re-analyses
MaterialIllust	additional material (information on additional material related to the dataset) for illustrations
MaterialOthers1	additional material (information on additional material related to the dataset) for other reuse methods 1
MaterialOthers2	additional material (information on additional material related to the dataset) for other reuse methods 2
MaterialOthers3	additional material (information on additional material related to the dataset) for other reuse methods 3
Hypotheses	hypotheses are usually provided
Sample	sample is usually provided
Design	research design is usually provided

Assignment	assignment is usually provided
PrimaryAnalyses	information on primary outcome analyses is usually provided
AdditionalAnalyses	information on additional outcome analyses is usually provided
MediationAnalyses	information on mediation analyses is usually provided
Publisher	publisher is usually provided
Year	year is usually provided
dupl1_Language	language is usually provided
Identifier	identifier is usually provided
Creator	creator (author name, identifier such as ORCID) is usually provided
Rights	rights (creative commons license valid for the object) are usually provided
Material	additional information on material associated with the dataset is usually provided
Hypotheses_3	hypotheses are very useful metadata
Hypotheses_2	hypotheses are slightly useful metadata
Hypotheses_1	hypotheses are no useful metadata
Sample_3	sample is very useful metadata
Sample_2	sample is slightly useful metadata
Sample_1	sample is no useful metadata
Design_3	research design is very useful metadata
Design_2	research design is slightly useful metadata
Design_1	research design is no useful metadata
Assignment_3	assignment is very useful metadata

Assignment_2	assignment is slightly useful metadata
Assignment_1	assignment is no useful metadata
PrimaryAnalyses_3	information on primary outcome analyses is very useful metadata
PrimaryAnalyses_2	information on primary outcome analyses is slightly useful metadata
PrimaryAnalyses_1	information on primary outcome analyses is no useful metadata
AdditionalAnalyses_3	information on additional outcome analyses is very useful metadata
AdditionalAnalyses_2	information on additional outcome analyses is slightly useful metadata
AdditionalAnalyses_1	information on additional outcome analyses is no useful metadata
MediationAnalyses_3	information on mediation analyses is very useful metadata
MediationAnalyses_2	information on mediation analyses is slightly useful metadata
MediationAnalyses_1	information on mediation analyses is no useful metadata
Publisher_3	publisher is very useful metadata
Publisher_2	publisher is slightly useful metadata
Publisher_1	publisher is no useful metadata
Year_3	year is very useful metadata
Year_2	year is slightly useful metadata
Year_1	year is no useful metadata
Language_3	language is very useful metadata

Language_2	language is slightly useful metadata
Language_1	language is no useful metadata
Identifier_3	identifier is very useful metadata
Identifier_2	identifier is slightly useful metadata
Identifier_1	identifier is no useful metadata
Creator_3	creator (author name, identifier such as ORCID) is very useful metadata
Creator_2	creator (author name, identifier such as ORCID) is slightly useful metadata
Creator_1	creator (author name, identifier such as ORCID) is no useful metadata
Rights_3	rights (Creative Commons license valid for the object) are very useful metadata
Rights_2	rights (Creative Commons license valid for the object) are slightly useful metadata
Rights_1	rights (Creative Commons license valid for the object) are no useful metadata
Material_3	additional information on material associated with the dataset is very useful metadata
Material_2	additional information on material associated with the dataset is slightly useful metadata
Material_1	additional information on material associated with the dataset is no useful metadata
hypotheses_use	perceived usefulness of hypotheses as metadata
sample_use	perceived usefulness of sample as metadata

assignment_use	perceived usefulness of assignment as metadata
primaryAnalyses_use	perceived usefulness of primary outcome analyses as metadata
additionalAnalyses_use	perceived usefulness of additional outcome analyses as metadata
mediationAnalyses_use	perceived usefulness of mediation analyses as metadata
publisher_use	perceived usefulness of publisher as metadata
year_use	perceived usefulness of year as metadata
language_use	perceived usefulness of language as metadata
identifier_use	perceived usefulness of identifier as metadata
creator_use	perceived usefulness of creator as metadata
rights_use	perceived usefulness of rights as metadata
material_use	perceived usefulness of information on additional material associated with the dataset as metadata
Sufficiency	is the annotation of own data with metadata sufficient
DDI	DDI documentation standard
DarwinCore	Darwin Core documentation standard
DublinCore	Dublin Core documentation standard
DataCiteMetadataSchema	Data Cite Metadata Schema
BIDS	BIDS documentation standard

StandardsOthers	other documentation standards
StandardsNone	documentation standards have not been used
MissingMetadata	metadata missing in existing documentation standards

---

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

Table 6

*Missing values*

Code	Label
-99	Unknown
-77	not specified (selection question)
-66	not specified (free text field)

---

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

## 8. Data analysis

### Sample description

After excluding the six test participants, the sample comprises 57 researchers (30 males and 27 females), with the majority coming from Germany (32 participants; for a complete overview see Table 7). The researchers belong to various psychological sub-disciplines, the most represented being social and cognitive psychologists (for a complete overview see Table 8). Frequent

research areas ranged from attention, emotions, and mental health to research methods (for all research areas see variable MainResearchArea in the dataset). Altogether, the sample composition might be slightly biased towards the previous research areas of the main investigator Dr. Katarina Blask (i.e. attention and attitudes/emotions).

Table 7

*Origin*

Origin	Frequency
Germany	32
France	2
Canada	2
Austria	10
The Netherlands	3
Ireland	1
USA	7

Table 8

*Sub-discipline*

Sub-discipline	Frequency
Biological Psychology	4
Clinical Psychology	4

---

Cognitive Psychology	13
Developmental Psychology	5
Differential Psychology	4
Educational Psychology	4
Health Psychology	7
Industrial Psychology	6
Social Psychology	9
Others	1

---

#### Primary data production relative to secondary data use

In order to determine the extent to which the surveyed researchers produce primary data relative to using secondary data, the according percentages (for means and standard deviations see Table 9) were tested against chance level (i.e. 50%). Both one-sample t-tests were statistically significant with  $t_{\text{primary}}(57) = 4.21, p < .001, d = 0.56$  and  $t_{\text{secondary}}(57) = -7.46, p < .001, d = 0.99$ . The larger effect size for secondary data use can be seen as an indicator of the continued under-representation of secondary data use in psychological research.



Table 9

*Means and standard deviations for primary data production and secondary data use*

	<i>M</i>	<i>SD</i>
Primary data production	69.72	35.38
Secondary data use	21.75	28.57

*Note.* M = Mean, SD = standard deviation.Frequent reuse purposes in Psychology

After having asked researchers to indicate the relative frequency with which they use secondary data compared to producing primary data, we asked them to indicate for which purposes they generally use data from their lab or other labs. The absolute frequencies for the different reuse purposes presented in Table 10 indicate that most researchers (almost 50%) reuse data from their own lab as well as from other labs for reanalysis. Moreover, there is a relatively high number of researchers using data from other labs for meta-analysis and systematic reviews (35%) compared to using their own data for these purposes (28% for meta-analysis and 21% for systematic reviews). For illustrations in lectures or on conferences, researchers preferably reuse their own data as opposed to those from other researchers (26% compared to 19%). Beyond these quite common reuse purposes, researchers also mentioned that they reuse data by aggregating multiple similar studies for a bigger sample, for power analysis or as a basis for simulation studies. The number of researchers who had never reused data neither from their own nor other labs was relatively small, with only nine researchers.

Table 10

*Frequent reuse purposes in Psychology*

Reuse purpose	Your lab	Other lab
Meta-analysis	16	20
Systematic review	12	20
Re-analysis	28	24
Illustrations	15	11
Others	9	5
None	13	10

*Note.* The cells include the absolute frequencies for the different reuse purposes.

Which metadata do researchers need for reuse?

In order to explore which metadata are most valuable for psychologists when reusing data for different purposes, we asked researchers which out of 12 method-specific metadata they would need to optimize their work in four pre-defined reuse scenarios (meta-analysis, systematic review, re-analysis, illustrations). Researchers could also name three further reuse purposes and indicate the corresponding metadata for them. The absolute frequencies for the selected metadata dependent on the corresponding reuse purpose are presented in Table 11. In virtue of the relatively small number of other reuse purposes mentioned by the researchers, the absolute frequencies of the metadata selected for these purposes are summarized.

Differences in the selected metadata depending on the reuse purpose were tested by conducting binary logistic regressions. Given the high diversity of the other reuse purposes indicated by the researchers and their small amount, we refrained from including them in the analyses. In

order to be able to regress the twelve different metadata on the four reuse purposes and to test for the simple contrasts between them, data were first restructured. That is, the four variables, including the absolute frequencies of a given metadata for the different reuse purposes, were restructured into four cases of the corresponding metadata that were dependent on the variable ReusePurpose (Meta-analyses vs. Systematic review vs. Re-analyses vs. Illustrations). In accordance with the descriptively apparent pattern displayed in Table 11, there were only significant differences between the metadata selected for illustrations compared to the other three reuse purposes (for a complete overview on all contrasts including p-values and Wald confidence intervals see Table 12). Not surprisingly, the descriptive analysis and the binary logistic regressions indicate that researchers need more method-specific metadata for a statistical reuse of psychological research data compared to a mere presentation purpose (e.g. illustrations in a lecture or at a conference).

Table 11

*Metadata needed for different reuse purposes*

Metadata	Meta-analysis	Systematic review	Re-analysis	Illustrations	Others
Hypotheses	23	25	25	3	6
Sample size	38	33	38	4	5
Inclusion/exclusion criteria	35	35	33	5	4
Setting	30	34	29	5	3
Sampling method	32	32	30	6	4

Kind of research design (e.g. correlational, experimental)	36	37	33	8	4
Included variables (e.g. dependent/independent variables)	36	36	37	9	5
Random/Non-random assignment	33	32	30	4	2
Primary outcome analyses	36	32	27	5	5
Additional outcome analyses	32	30	25	5	2
Mediation analyses	26	27	19	5	2
Material	25	25	32	7	3

*Note.* The cells include the absolute frequencies for the different metadata selected in the context of a specific reuse purpose.

Table 12

*Contrasts for reuse purposes dependent on the corresponding metadata*

Metadata	Contrast	<i>p</i>	Wald 95%-CI
Hypotheses	M – S	.702	[-.22, .15]

	M – R	.702	[-.22, .15]
	M – I	<.001	[.22, .50]
	S – R	1	[-.18, .18]
	S – I	<.001	[.25, .54]
	R – I	<.001	[.25, .54]
	M – S	.325	[-.09, .27]
	M – R	1	[-.17, .17]
	M – I	<.001	[.30, .63]
	S – R	.325	[-.27, .09]
	S – I	<.001	[.21, .54]
Samplesize	R – I	<.001	[.30, .63]
	M – S	1	[-.18, .18]
	M – R	.699	[-.15, .22]
Inclusion/exclusion criteria	M – I	<.001	[.39, .68]

Setting	S – R	.699	[-.15, .22]
	S – I	<.001	[.39, .68]
	R – I	<.001	[.35, .65]
	M – S	.444	[-.25, .11]
	M – R	.850	[-.17, .20]
	M – I	<.001	[.30, .60]
	S – R	.339	[-.09, .27]
	S – I	<.001	[.37, .67]
	R – I	<.001	[.28, .58]
Sampling method	M – S	1	[-.18, .18]
	M – R	.704	[-.15, .22]
	M – I	<.001	[.31, .62]
	S – R	.704	[-.15, .22]
	S – I	<.001	[.31, .62]

	R – I	<.001	[.27, .58]
Kind of research design (e.g. correlational, experimental)	M – S	.843	[-.19, .16]
	M – R	.559	[-.13, .23]
	M – I	<.001	[.34, .66]
	S – R	.434	[-.11, .25]
	S – I	<.001	[.36, .67]
	R – I	<.001	[.29, .60]
Included variables (e.g. dependent/independent variables)	M – S	1	[-.18, .18]
	M – R	.843	[-.19, .16]
	M – I	<.001	[.32, .64]
	S – R	.843	[-.19, .16]
	S – I	<.001	[.32, .64]
	R – I	<.001	[.34, .66]
	M – S	.848	[-.16, .20]

Random/Non-random assignment	M – R	.567	[-.13, .24]
	M – I	<.001	[.37, .66]
	S – R	.704	[-.15, .22]
	S – I	<.001	[.35, .65]
	R – I	<.001	[.32, .61]
Primary outcome analyses	M – S	.438	[-.11, .25]
	M – R	.082	[-.02, .34]
	M – I	<.001	[.41, .70]
	S – R	.342	[-.09, .27]
	S – I	<.001	[.33, .63]
	R – I	<.001	[.24, .54]
Additional outcome analyses	M – S	.704	[-.15, .22]
	M – R	.182	[-.06, .31]
	M – I	<.001	[.33, .63]



	S – R	.343	[-.10, .27]
	S – I	<.001	[.30, .60]
	R – I	<.001	[.21, .51]
Mediation analyses	M – S	.850	[-.20, .17]
	M – R	.174	[-.06, .31]
	M – I	<.001	[.22, .53]
	S – R	.120	[-.04, .32]
	S – I	<.001	[.24, .54]
	R – I	<.001	[.11, .39]
Material	M – S	1	[-.18, .18]
	M – R	.182	[-.31, .06]
	M – I	<.001	[.17, .48]
	S – R	.182	[-.31, .06]
	S – I	<.001	[.17, .48]
	R – I	<.001	[.29,.60]

---

*Note.* M = Meta-analyses; S = Systematic Review; R = Re-analyses; I = Illustrations, Wald 95%-CI = Wald 95%-confidence interval.

### Metadata usually provided during upload and their perceived usefulness

In order to get an idea of which metadata are currently provided by most researchers and as how useful they are perceived, we asked researcher to select from 14 pre-defined metadata those that they usually provide during upload in a data repository. Moreover, they had to indicate for each selected metadata on a three-point Likert scale (labeled “very useful metadata”, “slightly useful metadata”, and “no useful metadata”) as how useful they perceive this metadata. Table 13 shows an overview of the absolute frequencies and the corresponding usefulness ratings for the different metadata. Having a closer look at Table 13, it becomes apparent that researchers usually provide method-specific metadata, which may be due to the fact that, at least at the moment, research data are most often shared in the context of a research article publication. At the same time, these metadata are perceived as most useful for reusing the corresponding data. On the contrary, bibliographic metadata are less often provided, except for year and creator, and are also perceived as less useful.

Table 13

### *Metadata usually provided during upload and their perceived usefulness*

Metadata	absolute frequency	$M_{\text{usefulness}}$	$SD_{\text{usefulness}}$
Hypotheses	35	2.27	0.77
Sample	47	2.80	0.47
Research design	47	2.94	0.24

Assignment	44	2.79	0.41
Primary analyses	38	2.61	0.64
Additional analyses	29	2.41	0.68
Mediation analyses	18	2.06	0.75
Publisher	26	1.58	0.81
Year	57	2.03	0.93
Language	26	2.40	0.72
Identifier	22	2.31	0.86
Creator	36	2.63	0.65
Rights	21	2.38	0.78
Material	33	2.58	0.55

---

*Note.* M = mean, SD = standard deviation.

Finally, participants were asked whether they believe that the provided metadata are sufficient for reusing their data. Here it turned out that most researchers believe in the sufficiency of their documentation, as indicated by a one-sample t-test testing the sample mean against the scale midpoint (i.e., 2.5),  $t(55) = 8.14$ ,  $p < .001$ ,  $d = 1.1$ . This impression may be due to current publication practices in psychology, the most common being that journals prompt researchers to submit an empirical paper together with the corresponding data.

### Metadata standards already used in Psychology

Interestingly, but not surprisingly, almost all researchers ( $N = 50$ ) indicated that they have never used a certain documentation standard for annotating their data in the past. There were only five researchers who had previously used known standards like DDI ( $N = 2$ ), Dublin Core ( $N = 1$ ) or BIDS ( $N = 3$ ). Notwithstanding the nearly total absence of documentation standards in the annotation of psychological research data, researchers believe in the reusability of their research data. Whether this is really the case or not is an empirical question which should be investigated in future research. However, there is already some tentative evidence from our own research group that the present documentation of psychological research data is not sufficient in many cases (Blask, Jalynskij, & Gerhards, 2020).

## 9. References

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>.

## 10. Links

PsyCuraDat project website: [www.leibniz-psychology.org/institut/drittmittelprojekte/psycuradat](http://www.leibniz-psychology.org/institut/drittmittelprojekte/psycuradat)

Leibniz Institute for Psychology Information (ZPID): [www.leibniz-psychology.org](http://www.leibniz-psychology.org)

EFS questback: [www.questback.com](http://www.questback.com)

## Appendix

### A. Screenshot of the questionnaire realized in EFS questback (2019)

613456 Survey PsyCuraDat Teilnehmer können befehlen.

Seite Filter Dynamik Scoring Vorschau

Umfragesprache: "Englisch"

TITEL	ID	INFO	AKTIONEN
Welcome	3550652		
998 Welcome	7055473		
General information	3849418		
998 Project Description	7055560		
Demographic Data	3849714		
111 Sex	7151872		
141 Origin	7138487		
121 Subdiscipline	7056402		
441 Main research area	7151877		
Status Quo Data User	3550708		
143 Konstante Summe	7057809		
Status Quo Data User_1	3864259		
262 Re-use methods	7136240		
Re-use methods Others	3893546		
143 Re-use Methods others	7136621		
121 Data Types	7068539		
Status Quo Data User_2	3868210		
262 Metadata needed	7068563		
Data Provider	3893556		
998 Instruction Data Provider	7136179		
Status Quo Data Provider	3856413		
262 Metadata Upload	7068076		
111 Annotation sufficient	7068121		
Use of Existing Standards	3858035		
121 Metadata Standards	7071115		
Filter (v_133 = "1" or v_134 = "1" or v_135 = "1" or v_136 = "1" or v_137 = "1" or check_character_filter_any(text', v_138))	3858071		
Metadata Standards 1	3859413		
142 Metadata 1	7073242		
Endseite Beendet (31)	3849124		
998 Text und Bild	7074098		

Legende: Ausblendbedingung Pflichtfrage oder DAC Variablen Barrierefreier Fragetyp Frage als personenbezogenes Datum markiert

B. Request sent to the researchers, the EFPA, and the APA

B1. Request sent to the researchers English and German version

English version:

Dear researcher,

We write to kindly ask you to support our plan to develop a documentation standard for psychological research data and their underlying methods by participating in a short online survey. We (Leibniz Institute for Psychological Information and Documentation (ZPID)) have designed this survey as one part of the BMBF-funded project PsyCuraDat.

The goal of our project is to develop a documentation standard which is specific for psychological research data and their related methods. In essence, we want to figure out two things with the help of the online survey. On the one hand, we want to know from you as data user which information (i.e. metadata) you need for understanding and reusing an existing dataset. Thus, we want to determine the existing gaps in the documentation required for secondary data use from your perspective as a data user. On the other hand, we want to know which metadata you currently use to describe your data, and if you consider these metadata as sufficient for reusing your data. Uncovering existing gaps in the re-usability and provision of psychological research data, the survey is aimed at determining the concrete components that should be included in a documentation standard for psychological research data.

To take part in our survey (the deadline for participating is on January 12, 2020), please click [here](#).

Thank you for your support! If you'd like to be informed on the progress of our project, please send a short email with the subject "PsyCuraDat News" to [PsyCuraDat@leibniz-psychology.org](mailto:PsyCuraDat@leibniz-psychology.org)

Yours sincerely,

The PsyCuraDat project team



BMBF-Project PsyCuraDat 2.0: Online Survey (Metadata in Psychology)

Katarina Blask, Ph.D.

Lea Gerhards, M.A.

Erich Weichselgartner, Ph.D.

PS.: In case, we sent you this message in your function as a contact person for a certain researcher or a group of researchers, please forward this email to the respective person/s.

B2. Request sent to the European Federation of Psychologists' Association (EFPA):

Dear Ms. Steyaert,

with reference to your last email correspondence with Erich Weichselgartner regarding our project PsyCuraDat, I wanted to ask you whether you would support our project by sending the attached invitation letter to our online survey to all of your members.

In essence, we want to figure out two things with the help of the online survey. On the one hand, we want to know from researchers in psychology which information (i.e. metadata) they need for understanding and reusing an existing dataset. Thus, we want to determine the existing gaps in the documentation required for secondary data use from the perspective of data users. On the other hand, we want to know which metadata researchers currently use to describe their data, and if they consider these metadata as sufficient for reusing their data. Uncovering existing gaps in the re-usability and provision of psychological research data, the survey is aimed at determining the concrete components that should be included in a documentation standard for psychological research data.

To have a look at our survey, please click [here](#). If you have any questions or suggestions for improvement, please do not hesitate to tell us.

We look forward to your response and hope that you support us!

Yours sincerely,

The PsyCuraDat project team

Katarina Blask, Ph.D.

Lea Gerhards, M.A.

Erich Weichselgartner, Ph.D.

B3. Request sent to the American Psychological Association (APA)

Dear Ms. Aebersold,

We write to kindly ask you to support our plan to develop a documentation standard for psychological research data and their underlying methods by forwarding an invitation letter to your members. Within the BMBF-funded project PsyCuraDat, we (Leibniz Institute for Psychological Information and Documentation (ZPID)) have designed an online survey, which we would like to cordially invite all APA members to take part in.

The goal of our project is to develop a documentation standard which is specific for psychological research data and their related methods. In essence, we want to figure out two things with the help of the online survey. On the one hand, we want to know from researchers in psychology which information (i.e. metadata) they need for understanding and reusing an existing dataset. Thus, we want to determine the existing gaps in the documentation required for secondary data use from the perspective of data users. On the other hand, we want to know which metadata researchers currently use to describe their data, and if they consider these metadata as sufficient for reusing their data. Uncovering existing gaps in the re-usability and provision of psychological research data, the survey is aimed at determining the concrete components that should be included in a documentation standard for psychological research data.

Given that it is common practice in Psychology to describe research data within the frame of a research article, we want to build our standard on the JARS developed by your association. Building upon a well-established standard, we hope to decrease the perceived barrier for good data documentation as well as increase the comparability and compatibility of research articles and their underlying data.

To have a look at our survey, please click [here](#). If you have any questions or suggestions for improvement, please do not hesitate to tell us.

We look forward to your response and hope that you support our project!

Yours sincerely,

BMBF-Project PsyCuraDat 2.0: Online Survey (Metadata in Psychology)

The PsyCuraDat project team

Katarina Blask, Ph.D.

Lea Gerhards, M.A.

Erich Weichselgartner, Ph.D.