
PsyCuraDat: The development of a user-friendly curation standard for psychological research data

Katarina Blask, Marie-Luise Müller, Marc Latz, Valentin Arnold and Stephanie Kraffert

Leibniz-Institute for Psychology

The project *PsyCuraDat* aims to develop a user-friendly curation standard for psychological research data based on method-specific curation criteria, considering the needs of researchers in their roles as data providers and data users. This standard will bring enhanced effectiveness and efficiency, as well as increased satisfaction and quality to the curation and reuse of data in the field of psychology.

1 Introduction

The Open Science movement has brought useful knowledge and important changes to the scientific communities. Also within psychological science, the guidelines of open science practices are increasingly being implemented, in order to make research data openly accessible, and to enable sustainable (re-)use of data. However, standards specifically aiming at psychologists from all sub-disciplines, and allowing them to optimally prepare their data for reuse, hardly exist. To counteract this problem, we have started the project *PsyCuraDat*, to develop a user-friendly curation standard for psychological research data, meeting all necessary requirements to assure the data's long-term interpretability and reusability.

2 Approach & Development

In order to develop a curation standard, guaranteeing the sustainable use of psychological research data while meeting the needs of researchers in their role as data users and data providers, we began our project with a detailed examination of existing curation criteria and quality standards in psychology. We assessed whether existing standards are suitable for the documentation of all central phases of the research process (i.e., conceptualization, data collection, data analysis, publication and archiving). First results showed that none

Das hier beschriebene Poster ist in der Open Access-Plattform der Universität Heidelberg heiDOK unter der DOI: <https://doi.org/10.11588/heidok.00029714> veröffentlicht.

of the existing standards could adequately provide the documentation of the whole psychological research process. Also, existing standards are barely known, and thus barely used. Our research suggests that this is mostly due to too complicated handling [2]. Subsequently, we explored which data, and metadata, are needed for which types of reuse. We found that researchers primarily depend on information about the research design, hypotheses, a codebook and a study protocol [2]. Proceeding from those results, we elaborated method-specific curation criteria for a documentation standard, and developed a first prototype. Those curation criteria have led to a contextual specification, divided into three documentation levels based on the psychological research process: The first level represents the research objective on a *conceptual level*, and consists of the hypotheses, and the research design, as well as information on the sample. On the second level, the research design is presented on an *operational level*, on which researchers describe the methods used to operationalize the variables (e.g., through an extended codebook). The third level offers a detailed *description* of the research process, including a procedure graphic, as well as a presentation of the data preparation and analysis steps.

In order to test the usability of the prototype, two user studies were conducted with samples composed of psychological researchers. Following researchers' direct interaction with datasets, prepared in accordance with the standard, we used cognitive interviews to investigate the standard's usability. We primarily focused on evaluating the formal specification, as well as exploring which metadata must be provided, and in what form it should be presented. Our findings in user study 1 suggested that participating researchers perceived the contextual specification and the three documentation levels as rather technical, and not interlinked enough. Thus, we refined the standard towards a more comprehensible structure, which was then tested in user study 2. Besides improving the standard's contextual structure, additional data preparation and analysis scripts, as well as a short manual, describing the different documentation parts and their functional properties, were added [3]. Furthermore, results of both user studies showed that researchers most strongly relied on the conceptual description of the research design, the codebook, the analysis script and the graphical description of the procedure. Moreover, they requested a more hierarchical documentation structure of the procedure graphic (i.e., the graphic presented as a flow chart). Lastly, all participating researchers stated that they could imagine integrating the standard into their research process, and that they perceived the benefits of using this documentation standard to outweigh the costs and efforts [4].

3 Conclusions & Outlook

This empirically developed documentation standard has been based on curation criteria created in collaboration with psychological researchers from various sub-disciplines. Its content specification reflects the psychological research process and enables a detailed documentation of all central phases by providing a three-level-structure. This structure allows for a detailed description of the data on a study level (i.e., the conceptual de-

scription of the research design) as well as on a data level (i.e., the codebook) which, according to our empirical results, is both essential in order to fully understand and reuse a dataset. Furthermore, information provided on these two levels have to be linked to the data, and to each other, through a comprehensive description of the data collection and analysis process, including a summarizing procedure graphic, as well as all materials used during the data collection and analysis process. Besides developing a user manual and guidelines about how to implement the standard in everyday research routine, we are currently working out the standard's information architecture, aiming at a high level of comprehensibility and usability. That is to say, an "easy-to-use" and "easy-to-learn" data documentation structure, enabling effective and efficient use and reuse of psychological research data. When completed, our standard aims to offer a user-friendly tool to allow psychologists from all sub-disciplines to optimally prepare their data for reuse. It offers a contribution to Open Science and to the sustainable use of research data.

Acknowledgements

The project *PsyCuraDat* has been funded by the German Federal Ministry of Education and Research (funding number: 16QK08).

Bibliography

- [1] Blask, K., Gerhards, L., & Jalynskij, M. (2020). 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. ZPID (Leibniz Institute for Psychology). <https://doi.org/10.23668/PSYCHARCHIVES.2757>
- [2] Blask, K., Jalynskij, M., & Gerhards, L. (2020). Metadata in Psychology 1.0: What researchers really need - Study description of the data referring to the expert interviews conducted in the BMBF-funded project PsyCuraDat. ZPID (Leibniz Institute for Psychology). <https://doi.org/10.23668/PSYCHARCHIVES.2750>
- [3] Blask, K., Müller, M.-L., Arnold, V., & Kraffert, S. (2020). Evaluation of the PsyCuraDat- Specification 1.0 - Study description of the data referring to the first user study conducted in the BMBF-funded project PsyCuraDat. ZPID (Leibniz Institute for Psychology). <https://doi.org/10.23668/PSYCHARCHIVES.4318>
- [4] Blask, K., Müller, M.-L., Arnold, V., & Kraffert, S. (2021). Evaluation of the PsyCuraDat- Specification 2.0: Study description of the data referring to the second user study conducted in the BMBF-funded project PsyCuraDat. ZPID (Leibniz Institute for Psychology). <https://doi.org/10.23668/PSYCHARCHIVES.4459>