

Inhaltsverzeichnis

Replicating research

About the terminology

Repeatability requirements for text-oriented digital methods

2

2

3



This page was translated from German into English using **DeepL**.

Please help completing the translation.

(remove this paragraph once the translation is finished)

Replicating research

About the terminology

A much-discussed and current problem in many scientific disciplines is the traceability of research. A 2016 survey by the scientific journal Nature (Baker 2016) found that of about 1500 researchers, more than half reported having failed to reproduce **their own experiments**. Overall, 52% of respondents perceived a „significant 'crisis' of reproducibility.“ Low reproducibility rates have also been reported in other research fields. The largest attempt to date to replicate psychological studies concluded that of the 100 prominent papers examined, only 39% could be clearly replicated (Open Science Collaboration 2015). When attempting to replicate as accurately as possible authoritative experiments from basic cancer research described in «Nature», «Cell», «Science», and other respected science journals, in 54 of 112, the results when replicated were approximately the same as in the original experiment (Errington et al. 2021). However, in the majority of experiments, the results could not be confirmed. The paradigm shift from dominantly hermeneutic to empirical methods also confronts the (digitally working) humanities with new tasks of ensuring their own connectivity to established concepts, questions, and cognitive goals (cf. Schöch 2017). In the multidimensional dependencies that a repetitive research has on its original study, various definitions can be found in the research literature. Repeatedly mentioned are: Replication, reproduction, and reanalysis (Gómez et al. 2010, Hüffmeier et al. 2016). As a proxy for the conceptual clarification of the network of relations between original and repetition, three aspects are combined below:

- the questioning,
- the data, and
- the analytical methods,

and the resulting conceptualizations mentioned.

Schöch 2017, Fig. 1	Question		Data		Method	
	same	different	same	different	same	different
Replication (of the experiment)	x		x		x	
Reanalysis (of the data)	x		x			x
Reproduction (of the results)	x			x	x	
Follow-up research (to the question)	x			x		x
Reinterpretation (of the results)		x	x		x	
Reuse (of the data)		x	x			x

Schöch 2017, Fig. 1	Question		Data		Method	
	same	different	same	different	same	different
Reuse (of the code)		x		x	x	
N/A (no reference)		x		x		x

This form of typology simply describes the relationships between a study and its replication and is not intended to make a distinction in a purely binary way, because data or methods will rarely be completely identical or completely different. The term replication is used in this conceptual outline to refer to the exact repetition of a study. The same research question is addressed again using the same data and the same methods. The latter aspect can lead to slightly different methods of analysis and to a division into direct and conceptual replication:

- direct replication: an experiment is repeated under the same conditions.
- conceptual replication: the experimental conditions of the previous experiment are modified.

Replicability as a requirement is always spoken of when, in a quantitatively operating, scientific approach, the same results are also found in an investigation under the same conditions and using the same method.

References

- Baker, M. 1,500 scientists lift the lid on reproducibility. Nature 533, 452–454 (2016). <https://doi.org/10.1038/533452a>
- Camerer, Colin F. et al. (2016): „Evaluating replicability of laboratory experiments in economics“, in: Science 351.6280: 1433–1436, <https://doi.org/10.1126/science.aaf0918>.
- Errington, Timothy M. et al. (2021) Investigating the replicability of preclinical cancer biology eLife 10:e71601, <https://doi.org/10.7554/eLife.71601>.
- Gómez, Omar S. & Juristo, Natalia & Vegas, Sira. (2010). Replication, Reproduction and Re-analysis: Three ways for verifying experimental findings. 1st International Workshop on Replication in Empirical Software Engineering Research (RESER'2010).
- Hüffmeier, Joachim & Mazei, Jens & Schultze, Thomas. Reconceptualizing replication as a sequence of different studies: A replication typology, Journal of Experimental Social Psychology, Volume 66, 2016, Pages 81-92, ISSN 0022-1031, <https://doi.org/10.1016/j.jesp.2015.09.009>.
- Open Science Collaboration (2015), Estimating the reproducibility of psychological science, Science 349.6251, <https://doi.org/10.1126/science.aac4716>.
- Schöch, Christof. 2017. “Wiederholende Forschung in den digitalen Geisteswissenschaften.” In Konferenzabstracts DHd2017: Digitale Nachhaltigkeit, edited by DHd-Verband. <https://doi.org/10.5281/zenodo.277113>.
- Schöch, Christof & van Dalen-Oskam, Karin & Antoniak, Maria & Jannidis, Fotis & Mimno, David. (2020, June 14). Replication and Computational Literary Studies. Digital Humanities Conference 2020 (DH2020), Ottawa, Canada. <https://doi.org/10.5281/zenodo.3893428>.

Repeatability requirements for text-oriented digital methods

- **Availability of the starting materials** : The digitized corpus should be available in a common, open format (e.g. XML). In this context, open means not only free of technical or legal restrictions, but also freely accessible in the sense of Open Access. This also applies to the data

material that accompanies the corpus for processing, such as lists with

- configuration variables
 - exclusion or inclusion words
 - transliterations
 - principal part reduction
 - translations
- **Availability of the used tools** (Ideally all open source): In a broader sense, all applications that were used to process the source material. In the more restricted sense, it includes the corresponding tools in their respective development version.
 - operating systems
 - programming languages
 - databases
 - digital working environments (online and offline)
 - tool sets
 - Custom programming
 - **Availability extended data material** (metadata, documentation, literature): All materials necessary in classifying the source material, systematizing the research data, and assessing the results of the analysis should be freely accessible.
 - **Availability of the analysis data**: All data modified or newly created by the tools used should be freely and openly accessible.
 - **Availability of documentation of used procedures (algorithms) and data formats**: It is fundamentally important for evaluation and comprehensibility that both the individual process steps (in the case of text processing, for example, these could be: Normalization, principal part reduction, tokenization, transliteration, chunk splitting) as well as the mathematical or statistical relationships on which they are based. The analysis results and evaluations, if they were generated automatically from the process steps, should be seen as part of the research data, and the data formats used in the process should be documented.
 - **Long-term availability of research data and tools**: In the case of direct replication, the repetition of an experiment under exactly the same conditions of the previous experiment, it seems inevitable for reproducibility to archive not only the research data itself, but also the parts of the tools that are either not available in publicly accessible repositories, or that run the risk of no longer being available in the future in the version used due to version diversity during development.

From:

<http://replicatio.science/dokuwiki/> - **documentatio replicationis**

Permanent link:

<http://replicatio.science/dokuwiki/doku.php/en/glossar/replicatio>

Last update: **2023-03-16**

