Inhaltsverzeichnis

Wiederholende Forschung 2
Zu den Begriffsbildungen 2
Anforderungen an die Wiederholbarkeit bei textorientierten digitalen Methoden 3

Wiederholende Forschung

Zu den Begriffsbildungen

Ein vieldiskutiertes und aktuelles Problem in vielen Wissenschaftsdisziplinen ist die Nachvollziehbarkeit von Forschungsarbeiten. Eine Umfrage der naturwissenschaftlichen Fachzeitschrift Nature im Jahre 2016 (Baker 2016) ergab, dass von den rund 1500 Forscherinnen und Forschern über die Hälfte angab, daran gescheitert zu sein, die **eigenen Experimente** zu reproduzieren. Insgesamt 52% der Befragten nahmen eine „significant 'crisis' of reproducibility“ wahr. Auch in anderen Forschungsfeldern wird von niedrigen Reproduzierbarkeitsquoten berichtet. Der bis zu dem Zeitpunkt größte Versuch, psychologische Studien zu replizieren, kam zu dem Schluss, dass von den 100 untersuchten prominenten Arbeiten nur 39 % eindeutig repliziert werden konnten (Open Science Collaboration 2015). Bei dem Versuch, massgebliche Experimente aus der Krebsgrundlagenforschung, die in «Nature», «Cell», «Science» und weiteren angesehenen Wissenschaftsmagazinen beschrieben wurden, so exakt wie möglich zu wiederholen, fielen bei 54 von 112 die Resultate bei der Wiederholung ungefähr so aus, wie im Originalversuch (Errington et al. 2021). Jedoch bei der Mehrheit der Experimente ließen sich die Ergebnisse nicht bestätigen.

Der Paradigmenwechsel von dominant hermeneutischen hin zu empirischen Methoden stelle auch die (digital arbeitenden) Geisteswissenschaften vor neue Aufgaben, die eigene Anschlussfähigkeit an etablierte Konzepte, Fragestellungen und Erkenntnisziele sicherzustellen (vgl. Schöch 2017).

Bei den mehrdimensionalen Abhängigkeiten, die eine wiederholende Forschung zu ihrer Originalstudie besitzt, sind verschiedene Begriffsbestimmungen in der Forschungsliteratur zu finden. Immer wieder genannt werden: Replikation, Reproduktion und Reanalyse (Gómez et al. 2010, Hüffmeier et al. 2016). Stellvertretend für die konzeptionelle Klärung des Beziehungsgeflechtes zwischen Original und Wiederholung werden nachfolgend drei Aspekte kombiniert:

- die Fragestellung,
- die Daten und
- die Analysemethoden,

und die sich daraus ergebende Begriffsbildungen genannt.

Schöch 2017, Abb. 1	**Frage**		**Daten**		**Methode**	
	gleich	**anders**	**gleich**	**anders**	**gleich**	**anders**
Replikation (des Experimentes)	x		x		x	
Reanalyse (der Daten)	x		x			x
Reproduktion (der Ergebnisse)	x			x	x	
Nachfolgeforschung (zur Fragestellung)	x			x		x
Reinterpretation (der Ergebnisse)		x	x		x	
Nachnutzung (der Daten)		x	x			x

Schöch 2017, Abb. 1	Frage		Daten		Methode	
	gleich	anders	gleich	anders	gleich	anders
Nachnutzung (des Codes)		x		x	x	
N/A (kein Bezug)		x		x		x

Diese Form der Typologie beschreibt lediglich die Beziehungen zwischen einer Studie und ihrer Wiederholung und ist nicht dazu gedacht, eine Unterscheidung in rein binärer Weise vorzunehmen, denn Daten oder Methoden werden in den seltensten Fällen komplett identisch oder völlig verschieden sein. Der Begriff Replikation wird in dieser konzeptuellen Gliederung benutzt, die exakte Wiederholung einer Studie zu bezeichnen. Die gleiche Forschungsfrage wird auf gleicher Datenlage mit den gleichen Methoden erneut bearbeitet. Wobei es bei letzterem Aspekt zu leicht abweichenden Analysemethoden und in eine Einteilung in direkte und konzeptionelle Replikation kommen kann:

- direkte Replikation: Ein Experiment wird unter den gleichen Bedingungen wiederholt.
- konzeptionelle Replikation: Die Versuchsbedingungen des Vorgängerversuchs werden modifiziert.

Von Replizierbarkeit als Forderung wird immer dann gesprochen, wenn in einem quantitativ operierendem, wissenschaftlichen Ansatz bei einer Untersuchung unter den selben Bedingungen und unter Anwendung der derselben Methode auch dieselben Ergebnisse gefunden werden.

Literatur

- 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.

Anforderungen an die Wiederholbarkeit bei textorientierten digitalen Methoden

- **Verfügbarkeit der Ausgangsmaterialen**: Das digitalisierte Korpus sollte in einem gängigen,

offenen Format (z.B XML) vorliegen. Offen bedeutet in diesem Zusammenhang nicht nur frei von technischen oder rechtlichen Einschränkungen, sondern auch frei zugänglich im Sinne von Open Access. Dies betrifft auch das Datenmaterial, was begleitend zum Korpus für die Verarbeitung benutzt wird, wie z.B. Listen mit
    - Konfigurationsvariablen
    - Exklusions- oder Inklusionswörtern
    - Transliterationen
    - Stammformreduktionen
    - Übersetzungen
- **Verfügbarkeit der eingesetzten Werkzeuge** (Idealer Weise alle als Open Source): Im weiteren Sinne alle Anwendungen, die genutzt wurden, um das Ausgangsmaterial zu bearbeiten. Im engeren Sinne umfasst es die entsprechenden Werkzeuge in ihrer jeweiligen Entwicklungsversion.
    - Betriebssysteme
    - Programmiersprachen
    - Datenbanken
    - Digitale Arbeitsumgebungen (On- und Offline)
    - Tool-Sets
    - Eigenprogrammierungen
- **Verfügbarkeit erweitertes Datenmaterial** (Metadaten, Dokumentationen, Literatur): Alle Materialen, die bei der Einordnung des Ausgangsmaterials, Systematisierung der Forschungsdaten und Einschätzung der Analyseergebnisse notwendig sind, sollten frei zugänglich sein.
- **Verfügbarkeit der Analysedaten**: Alle Daten, die durch die eingesetzten Werkzeuge verändert oder neu erzeugt wurden, sollten frei und offen zugänglich sein.
- **Verfügbarkeit der Dokumentation von eingesetzten Verfahren (Algorithmen) und Datenformaten**: Für die Bewertung und Nachvollziehbarkeit fundamental wichtig ist, dass sowohl die einzelnen Verfahrensschritte (Bei Verarbeitung von Text könnten dies beispielsweise sein: Normalisierung, Stammformreduktion, Tokenisierung, Transliteration, Chunk-Splitting) als auch die ihnen zugrunde liegenden mathematischen oder statistischen Zusammenhänge aufgezeigt werden. Die Analyseergebnisse und -auswertungen sollten, sofern sie aus den Verfahrensschritten automatisch erzeugt wurden, als Bestandteil der Forschungsdaten gesehen, die dabei eingesetzten Datenformate dokumentiert werden.
- **Langzeitverfügbarkeit von Forschungsdaten und Werkzeugen**: Bei direkter Replikation, der Wiederholungen eines Experimenten unter exakt den gleichen Bedingungen des Vorgängerversuchs, scheint es für die Reproduzierbarkeit unumgänglich, nicht nur die Forschungsdaten selbst zu archivieren, sondern auch die Teile der Werkzeuge, die entweder nicht in öffentlich zugänglichen Repositorien verfügbar sind, oder die aufgrund von Versionsvielfalt bei der Entwicklung Gefahr laufen, in der benutzten Version in der Zukunft nicht mehr zur Verfügung zu stehen.

From:
http://replicatio.science/dokuwiki/ - **documentatio replicationis**

Permanent link:
http://replicatio.science/dokuwiki/doku.php/de/glossar/replicatio

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