

Transparente Wissenschaft als eine Voraussetzung für Replizierbarkeit

Erich Weichselgartner

Leibniz-Zentrum für Psychologische
Information und Dokumentation,
ZPID, Trier

Gefördert von



DFG-Rundgespräch *Replizierbarkeit und Replikation*, Bonn, 15.12.2015

1. PsychData

Die fünf Säulen von PsychData

Drei Säulen aus Ursprungsprojekt

1. (Langzeit-) Archivierung

- Keine proprietären Formate (ASCII), laufende IT-Migration
- Selektion in Bezug auf erwartete Nachnutzung:

den und hieran orientiert die Datenauswahl vornehmen. Dokumentiert und archiviert werden sollten vorerst nur die Daten, bei denen von einem realistischen Nachnutzungsinteresse ausgegangen werden könne. Hierzu zählten vor allem Längsschnittuntersuchungen, große Umfragestudien einschließlich historisch einmaliger Studien- insbesondere aus den Bereich der Klinischen, der Entwicklungs-, der Pädagogischen, der Geronto- sowie der Arbeits- und Organisationspsychologie. Das Projekt sollte daher vornehmlich auf Datenmaterial aus diesen Bereichen begrenzt werden.

1. PsychData

Die fünf Säulen von PsychData

Drei Säulen aus Ursprungsprojekt

1. (Langzeit-) Archivierung
 - Keine proprietären Formate (ASCII), laufende IT-Migration
 - Selektion in Bezug auf erwartete Nachnutzung:
2. Dokumentation
 - Forschungskontext (DC, DDI), Variablenebene (Kodebuch)
3. Bereitstellung
 - Wissenschaftliche Nutzung, Vertrag, revisionssichere CD-ROM, kostenlos

Data Sharing: Die Veröffentlichung von Forschungsdaten zur Nachnutzung durch andere.

Vorteile (nicht nur für die Psychologie):

- Umfassende Auswertung
- Vermeidung redundanter Datenerhebungen
- Bereitstellung nicht replizierbarer Daten
- Analyse unter neuen Fragestellungen und Perspektiven
- Reanalyse mit anderen Analysemethoden
- vergleichende Analyse
- Aggregation von Daten
- Ermittlung historischer Veränderungen
- Erfüllung der Grundsätze zur Sicherung guter wissenschaftlicher Praxis, usw. (Weichselgartner, 2008).

Replikation ist nicht Hauptziel des *Data Sharing*.

Die fünf Säulen von PsychData

Drei Säulen (Ursprungsprojekt)

1. (Langzeit-) Archivierung
 - Keine proprietären Formate (ASCII), laufende IT-Migration
2. Dokumentation
 - Forschungskontext (DC, DDI), Variablenebene (Kodebuch)
3. Bereitstellung
 - Wissenschaftliche Nutzung, Vertrag, revisionssichere CD-ROM, kostenlos

Vierte Säule (DataWiz)

4. Forschungsbegleitende Assistenz, Autonomie

- 1. ***Für welche Zielgruppe ist die Plattform geeignet?***
 - **Datengeber:** Treuhändische Langzeitarchivierung nach **disziplinspezifischen** Gesichtspunkten; Selektionskriterien finden Anwendung! Explizite Nutzungsregelungen (z.B. Karenzzeiten). Quantitative empirische Forschung.
 - **Datennehmer:** Forschung und Lehre in Wissenschaft.

2. *Wie wird PsychData angenommen / genutzt?*

Umfang

157 Datensätze mit > 36 Millionen Datenpunkten

Nutzung (2015)

6 Nachnutzungen von Datensätzen

[insgesamt: universitäre (45%), außerunivers. Forschung (36%); Lehre; interdisziplinär: Sportwiss., Robotik, Linguistik, Erziehungswiss.]

F : L = 6 : 3

19 Nutzerberatungen

13329 Besucher der Webseiten

Handbuch und Verträge

Rückmeldung

(Retrospektive) Dokumentation zu aufwendig

Kontrolle über Daten behalten

→ DataWiz

3. Welche Aspekte halten Forscher ggf. davon ab ihre Daten hier (und auf anderen Plattformen) zu teilen?

Ethische Gesichtspunkte

1. Datenschutz, Vertraulichkeit

Regulatorische Gesichtspunkte

1. *Biopsychologie: „Wir generieren im Jahr 3 TB Daten. Wir wollen den Regeln guter wiss. Praxis genügen und die Daten aufbewahren. Bei Bedarf können **wir** die archivierten Daten verstehen“.*

Ökonomische Gesichtspunkte

1. *Wir haben einen riesigen Aufwand und mehr als eine Million Euro in das Projekt gesteckt. Andere sollen jetzt kostenfrei davon profitieren?*
2. *„Im SFB (Federführung Psychologie) scheint sich gerade die Tendenz durchzusetzen, eine offene, **unstrukturierte** Plattform wie z.B. Zenodo oder DataDryad zu nutzen.“*

Welche Aspekte halten Forscher ggf. davon ab ihre Daten hier (und auf anderen Plattformen) zu teilen?

Niederschwellige Repositorien

Keine formalen Qualitätsanforderungen, -prüfungen und Dokumentationsstandards: ReadMe-Files, Study Materials; Hinweis: „*See original article for references on how data was collected and analysed.*”

1. DataDryad

Psychological Science (APS); “Data are provided according to community standards”

2. Dataverse

3. Figshare

PLoS

4. Zenodo

5. OSF

6. Interuniversity Consortium for Political and Social Research (ICPSR)

Archives of Scientific Psychology Data (APA)

Gemischte Standards: Teilweis hohe Qualität entsprechend „[Guide to Social Science Data Preparation and Archiving, 5th Edition](#)”

*Tabellarische Übersicht
verfügbar*

Welche Aspekte halten Forscher ggf. davon ab ihre Daten hier (und auf anderen Plattformen) zu teilen?

Nicht-Kommunikation negativer Ergebnisse

‘Negative results’ now account for only 14% of published papers, down from 30% in 1990.
([How science goes wrong](#), The Economist, 19.10.2013)

Psychologie

An Archive of Brief Reports of Replication Attempts in Experimental Psychology.

<http://psychfiledrawer.org/>

Systemkritik

Erhöhung von Transparenz, Verbesserung der Reproduzierbarkeit⁺
(Weichselgartner: *Towards Enhanced Science in Psychology*, ECP 2013, Stockholm)

The Open Science Framework (OSF)

An online infrastructure for documenting, archiving, logging, sharing, and registering scientific projects.

The Reproducibility Project

An open, large-scale collaborative effort to systematically examine the rate and predictors of reproducibility in psychological science.

The Potsdam Mind Research Repository

Provide access to peer-reviewed publications along with data and scripts for analyses and figures reported in them.

PsychFileDrawer.org

A tool designed to address the *File Drawer Problem* as it pertains to psychological research: the distortion in the scientific literature that results from the failure to publish non-replications.

ZPID-Initiativen

PsychOpen, PsychData, PubPsych

⁺*Publication Bias, Reporting Bias*

2. Replizierbarkeit von Studien

Replizierbarkeit (*replicability*) vs. Reproduzierbarkeit (*reproducibility*)

- Replicability. “The replication of scientific findings using **independent investigators, methods, data, equipment, and protocols** has long been, and will continue to be, the standard by which scientific claims are evaluated. However, in many fields of study there are examples of scientific investigations that cannot be fully replicated because of a lack of time or resources.”
- “Reproducible research” requires that data sets and computer code be made available to others for verifying published results and **conducting alternative analyses**. (Peng, 2009)

Thus “reproducibility” means that if we **start from the data gathered by the scientist** we can reproduce the same results, p-values, confidence intervals, tables and figures as those reported by the scientist.



Dagstuhl Perspectives Workshop:

“Data availability *per se* does not guarantee either reproducibility nor interoperability; reporting structures, minimal amounts of information and ontologies have been used in other domains to facilitate understanding experimental information. How could these be implemented in Psychology?”

Verknüpfung von Erkenntnissen innerhalb einer Disziplin und zwischen Disziplinen

- Berichtswesen standardisieren (*reporting standards, e.g. JARS*)
- Datenstandards und Ontologien etablieren
- Meta-Informationen (Annotationen):
 - fachliche Metadaten (DDI, caDSR, FITS)
 - bibliothekarische Metadaten (DC oder MODS haben bibliographischen Zweck!)

Ziel: “On the shoulders of giants”

Besonderheit der Psychologie: Untersuchung verändert
Untersuchungsgegenstand.

Ausweg Replizierbarkeitsproblematik: Validierung von Erkenntnissen
über Disziplinen hinweg. Erkenntnis wird aus mehreren Blickwinkeln
abgesichert: Neurochemie, Neurophysiologie, Pharmakologie,
Psychiatrie, Psychologie, etc. (Beispiel: DoD-Suizidforschung).

Standardisierung, Ontologien.

3. Registrierung von Studien

Pre-Registration: Beispiele aus der Psychologie

- Preregistered [Special Issue of the Journal of Media Psychology \(2017\)](#). Call for Papers
"The Special Issue is also dedicated to promote open science practices, requiring all authors to share all data and materials with the reviewers during the two evaluation stages, and with the wider community upon publication (as publicly available supplementary materials on the journal's website)."
- [Comprehensive Results in Social Psychology](#) is devoted to publishing social psychological research using the registered report format where a plan for the research is submitted for initial review. The journal aims to foster empirical debate, testing of theories, scientific discovery, and replications, all under the auspices of open scientific communication and exchange.
- [Cortex](#). Cortex is pleased to announce the launch of a new innovation in scientific publishing called the Registered Report. Different to established publishing models, Registered Reports divide the review process into two stages. Experimental methods and proposed analyses are pre-registered and reviewed before data are collected.

Deutsches Register klinischer Studien

https://drks-neu.uniklinik-freiburg.de/drks_web/

International

[ClinicalTrials.gov](https://clinicaltrials.gov)

(67 Treffer für Psychologie und Location = Germany)

Literatur

Weichselgartner, E. (2008). Fünf Jahre Primärdatenarchivierung in der Psychologie: Ein Erfahrungsbericht. In Ockenfeld, M. (Hrsg.), *Verfügbarkeit von Information* (S. 259-267). Frankfurt a. M.: DGI.

Kontakt

Erich Weichselgartner

Leibniz-Zentrum für Psychologische
Information und Dokumentation
(ZPID), Trier

Tel: +49 651 201-2056

E-Mail: wga@zpid.de

Web: <http://www.zpid.de/>