

## Open Science

Erich Weichselgartner\*

# DataWiz: Integration von Open-Science-Praktiken in den Forschungszyklus

DOI 10.1515/iwp-2017-0023

**Druck auf das Wissenschaftssystem.** In den letzten Jahren ist das etablierte Wissenschaftssystem von verschiedenen Seiten unter Druck geraten. Dabei handelt es sich zum einen um Kritik an der Art und Weise, wie Wissenschaft betrieben wird (z. B. Replikationskrise), und zum anderen um Kritik an der Art und Weise, wie wissenschaftliche Ergebnisse kommuniziert werden (z. B. Zeitschriftenkrise). Die Open-Science-Bewegung greift beide Aspekte auf und will hin zu einer transparenteren Wissenschaft in allen Phasen des Forschungszyklus. Der Forschungszyklus mit seiner idealisierten Vorstellung vom Ablauf des empirischen Forschungsprozesses von Ideenfindung über Untersuchungsplanung bis zur Kommunikation der Ergebnisse ist gut geeignet, die Rollen der beteiligten Akteure herauszuarbeiten. Open Science wendet sich zum einen an die Wissenschaftler selbst, die u. a. Forschungsdaten und Analysecode offenlegen, und zum anderen an Infrastruktureinrichtungen, die u. a. kosten- und barrierefreien Zugang zu bibliographische Daten und Fachaufsätzen schaffen sollen.

**Erhöhung der Transparenz.** Initiativen zur Offenlegung gibt es schon länger. Man denke an die Open-Source-Bewegung, die bereits in den 1970er Jahren mit Berkeley Unix begann. Oder an freie literarische Texte, die im Rahmen des Projekt Gutenberg ebenfalls seit den 1970er Jahren zur Verfügung stehen. Forschungsdaten werden von dem in die GESIS integrierten Zentralarchiv für Empirische Sozialforschung schon seit den 1960er Jahren zur Nachnutzung angeboten. Neuerdings werden diese und andere offengelegte Bausteine *wissenschaftlichen Arbeitens* unter dem Begriff Open Science subsumiert. Verschiedene Initiativen und Projekte sind engagiert dabei Open Science in die Wissenschaftsgemeinschaft einzuführen. Die Europäische Kommission bietet mit dem 2014 gestarteten Projekt FOSTER<sup>1</sup> (Facilitate Open Science Training for European

Research) eine Lernplattform für eine breite Vielfalt von Open-Science-Praktiken an: Open Source, Open Access, Open Data, Open Notebooks, Open Workflows, Open Repositories, Open Metrics und Open Review. In der Psychologie vergibt die Association for Psychological Science (APS) seit 2014 sog. *Open-Practice-Badges* für offenes wissenschaftliches Arbeiten, und zwar für *Open Data*, *Open Materials* und *Präregistrierung der Analyseverfahren*.<sup>2</sup> Übergeordnetes Ziel der Integration von Open Science in die täglichen Arbeitsabläufe von Forscherinnen und Forschern ist eine verantwortungsvoll betriebene Forschung. Das Open-Science-Training und die Vergabe der Open-Science-Badges soll eine verbesserte (transparentere) Darstellung wissenschaftlichen Arbeitens anregen.

**Auf den Schultern von Riesen?** Ursprünglich geht das Open-Practice-Plaketten-System auf das Center for Open Science an der University of Virginia (USA) zurück, das Anfang 2013 von den Psychologen Brian Nosek und Jeffrey Spies mit dem Ziel gegründet wurde, die Offenheit, Integrität und Reproduzierbarkeit wissenschaftlicher Forschung zu erhöhen. Um letztere scheint es nicht so gut bestellt zu sein, wie man angesichts der essentiellen Bedeutung von Replizierbarkeit für die moderne Erkenntnisgewinnung annehmen möchte. Zwar hat schon Bernhard von Chartres im Mittelalter mit seinem Gleichnis von den „Zwergen auf den Schultern von Riesen“<sup>3</sup> darauf hingewiesen, dass Fortschritt dadurch zustande kommt, dass Gelehrte dem vorgefundenen Wissensschatz ihren eigenen bescheidenen Beitrag hinzufügen und auch Leibniz meinte: „Es lohnt sich, die Entdeckungen anderer zu studieren, da für uns selbst eine neue Quelle für Ideen entspringt.“ Doch was ist, wenn die Riesen schwach und die Fundamente brüchig sind, weil publizierte Befunde nicht replizierbar sind (also nicht einmal ein kleiner Beitrag im Sinne von Chartres geleistet wird)?

**Wissenschaftskrise.** (s. dazu auch S. 154) In dem großangelegten „Reproducibility Project: Psychology“, in dem die Originalbefunde von 100 Veröffentlichungen in

<sup>1</sup> <https://www.fosteropenscience.eu/resources>.

\*Kontaktperson: PD Dr. Erich Weichselgartner, Leibniz-Zentrum für Psychologische Information und Dokumentation (ZPID), Universitätsring 15, 54296 Trier, E-Mail: [wga@zpid.de](mailto:wga@zpid.de)

<sup>2</sup> <https://www.psychologicalscience.org/publications/badges>.

<sup>3</sup> Wird von Google Scholar als Motto verwendet.

bedeutenden Psychologie-Zeitschriften repliziert werden sollten, haben Nosek und Kollegen, je nach Interpretation, weniger als 50 Prozent bestätigt gefunden (Baker, 2015). Dabei schneidet die Psychologie noch gut ab. In dem zweiten „Reproducibility Project: Cancer Biology“ zeichnen sich erheblich schlechtere Replikationsraten ab (Baker & Dolgin, 2017). Pritsker hat auf der Konferenz „Academic Publishing in Europe“ (2015) ausgeführt, dass die Replikationsrate im Wissenschaftsbetrieb allenfalls 30 Prozent sei (vgl. auch Pritsker, 2013). In der Medizin sorgte eine ganze Serie von Artikeln im Fachblatt *The Lancet* für Aufsehen, die in dem Fazit kulminierte, dass zu viel Forschungsmüll produziert werde (Macleod et al, 2014). Bereits 2009 waren Chalmers und Glasziou, ebenfalls im *Lancet*, zu der Schlussfolgerung gelangt, dass die falschen Forschungsfragen gestellt und mit mangelhaften Untersuchungsdesigns gearbeitet werde, und dass vor allem vorab nicht geprüft werde, ob vergleichbare Untersuchungen schon existieren (Chalmers & Glasziou, 2009). Letzteres ist wegen dem sog. Publication Bias gar nicht so leicht feststellbar, weil negative Resultate nach verbreiteter Praxis gar nicht zur Publikation angenommen werden (Rosenthal, 1979). Grund ist, dass die großen Fachzeitschriften den Fokus auf neue Befunde legen (Neophilia) oder, wie Nobelpreisträger Schekman es ausgedrückt hat, nach sexy Themen und herausfordernden Behauptungen lechzen (Schekman, 2013). Dieses Aufmerksamkeits-Heischen kann mitunter auch in handfeste Skandale münden. Der Stammzellenforscher Hwang Woo-suk wurde als Fälscher enttarnt und *Science* musste zwei Arbeiten zurückziehen. Der Sozialpsychologe Diederik Stapel hatte in Dutzenden von Arbeiten die Daten sogar frei erfunden und auch hier musste *Science* Publikationen widerrufen. Die aktuelle Entwicklung um Rücknahmen wissenschaftlicher Publikationen kann man im Retraction Watch Blog verfolgen.<sup>4</sup>

Generell führe der hohe Publikationsdruck zu fragwürdigen Forschungspraktiken, die von 35 Prozent der Befragten im Rahmen einer Metaanalyse sogar zugegeben wurden (Fanelli, 2009). Hinzu kommen weitere Formen der Tendenziosität, die sich aus der Forschungsförderung ergeben können. Wojick und Michaels (2015) sprechen von „Funding-induced Biases in Science“ und zählen 15 verschiedene Indikatoren auf. Neben voreingenommenen Gutachtern und Datenmanipulationen wird explizit auch die Weigerung genannt, Forschungsdaten weiterzugeben (zu teilen). Zwar muss dahinter nicht immer unlautere Absicht stecken, aber Fakt ist, dass Forschungsdaten nach wie vor schlecht zugänglich sind (vgl. Dehnhard, Weichselgartner

& Krampen, 2013; Vines et al, 2014; Wicherts et al, 2006). Dabei gibt es mannigfaltige Gründe, warum das Teilen von Forschungsdaten ausgesprochen wertvoll ist (Weichselgartner, 2008): Sie reichen von der Forschungsökonomie (bessere Nutzung von Ressourcen) und Forschungsethik (Vermeidung redundanter Datenerhebungen) hin zur Forschungslogik (Forschungsdaten sind ein wesentlicher Aspekt bei der Prüfung und Bewertung wissenschaftlicher Ergebnisse).

**Verbesserte Zugänglichkeit von Forschungsdaten.** Forschungspolitik und Forschungsförderer haben den Wert des *Data Sharing* vielleicht noch stärker erkannt als die Wissenschaftler selbst und es laufen seit Jahren intensive forschungspolitische Bemühungen zur verbesserten Zugänglichkeit. Beispielhaft seien die „Declaration on Access to Research Data from Public Funding“ der *Organisation for Economic Cooperation and Development* genannt (OECD, 2004), oder die Gründung der *Research Data Alliance* in 2013 mit Vertretern aus über 90 Ländern. Die Mitglieder der RDA wollen die internationale Dateninfrastruktur verbessern. Als nationale Forschungsförderer haben BMBF und DFG in den letzten Jahren viele einschlägige Fördermaßnahmen aufgelegt. Im Rahmen des DFG-Förderprogramms „Informationsinfrastrukturen für Forschungsdaten“ wurde auch das Projekt „DataWiz – Ein Assistenzsystem für das Management psychologischer Forschungsdaten“ bewilligt, das im Folgenden näher erläutert wird.

**DataWiz.** Fachwissenschaftler sind keine Experten für Datenmanagement. In den meisten Forschungsprojekten stehen solche Experten auch nicht zur Verfügung. Nicht nur deswegen wird das Management von Forschungsdaten<sup>5</sup> als unangenehme Zusatzbelastung empfunden. Diese Rückmeldung hat das ZPID im Zusammenhang mit dem Betrieb seines vom RatSWD akkreditierten Forschungsdatenzentrums (FDZ) PsychData immer wieder erhalten. Erschwerend kommt hinzu, dass für diese Zusatzbelastung kein unmittelbarer eigener Vorteil erkennbar ist. Zwar ändert sich dies langsam, weil sich herumspricht, dass das Teilen von Forschungsdaten die eigenen Zitationsraten erhöht (Piwowar et al, 2007), aber im Großen und Ganzen findet Forschungsdatenmanagement nach wie vor kaum statt. DataWiz setzt deshalb auf vier Anreize: (1) Auf die **Erleichterung** der Arbeit, indem es sich in den Workflow der Forschungspraxis integriert. Zum Beispiel enthält DataWiz ein Modul zum Überprüfen der Integrität von Daten. (2) Auf **Expertenwissen**, das kontextabhängig über

<sup>4</sup> <http://retractionwatch.com/>.

<sup>5</sup> Die Forschungsplanung sowie die Aufbereitung, Beschreibung, Archivierung und Bereitstellung der Daten.

eine breit angelegte Wissensbasis abrufbar ist. In der Phase der Versuchsplanung wird zum Beispiel auf die Ethiknorm hingewiesen, dass Einverständniserklärungen von Versuchspersonen einzuholen sind, und bei Bedarf können Mustererklärungen abgerufen werden. (3) Auf die Gewährleistung der nachhaltigen **Sicherung** der Daten. DataWiz unterstützt beispielsweise die Umwandlung proprietärer Statistikformate in Formate für die langfristige Aufbewahrung (JSON, PDF/A). (4) Auf **Autonomie**, denn DataWiz kann völlig selbständig und autark betrieben und konfiguriert werden. Neben dem befürchteten hohen Arbeitsaufwand im Datenmanagement haben Forscher nämlich gegenüber dem FDZ PsychData immer wieder die Sorge geäußert, sie würden die Kontrolle über die eigenen Daten verlieren. Aus diesem Grund ist ein Kernfeature von DataWiz ein fein abgestimmtes Rechtemanagement, das Nutzern erlaubt zu entscheiden, wer die Daten einsehen darf und ob und an welches Datenarchiv sie übergeben werden. Auf ausdrücklichen Wunsch der Fachgemeinschaft können dabei die Strukturen von Lehrstühlen oder Arbeitsgruppen abgebildet werden. Dadurch wird der gemeinsame Zugriff auf Daten in Forschergruppen oder bei der Betreuung von Qualifikationsarbeiten möglich.

**Integriertes Data-Lifecycle-Management.** Neben den genannten Aspekten trägt DataWiz auch dadurch zur Qualitätssicherung von Forschungsdaten bei, dass es schon zu Beginn des Forschungszyklus eingesetzt wird. Es hat sich immer wieder gezeigt, dass die Auseinandersetzung mit den Erfordernissen der Datenarchivierung und -veröffentlichung nicht erst am Ende des Forschungsprozesses erfolgen sollte. Nicht selten lassen sich dann nämlich wichtige Aspekte der Datenerhebung nicht mehr rekonstruieren, beteiligte Forscher haben die Stelle gewechselt oder es wurde vergessen, sich die für die Veröffentlichung der Daten erforderlichen Rechte zu besorgen (vgl. Weichselgartner, Günther & Dehnhard, 2011). DataWiz bietet ein in den Forschungsprozess integriertes Data-Lifecycle-Management. Bereits während der Konzeptions- und Durchführungsphase eines Forschungsprojektes werden die Dokumentation, Archivierung und Veröffentlichung von Forschungsdaten vorbereitet, indem beispielsweise die für die Dokumentation erforderlichen Metadaten (manuell, semi- oder vollautomatisch) erfasst und standardkonform aufbereitet werden. Daten werden dann dokumentiert, wenn sie anfallen und Informationen müssen nicht mühsam im Nachhinein rekonstruiert werden. Zum Beispiel werden die Informationen aus Datenmanagementplänen, die man mit DataWiz erstellt, automatisch in die Studiendokumentation übernommen. Doppelarbeit entfällt. Die Nutzer von DataWiz erhalten prozedurale Unterstützung, die in den wissenschaftlichen

Arbeitsprozess integriert ist und dazu beiträgt, den häufig als überflüssig erlebten Aufwand des Datenmanagements soweit wie möglich zu minimieren, behalten dabei aber die Kontrolle über ihre Forschungsdaten.

**Technische Umsetzung.** Technisch betrachtet enthält DataWiz zwei Komponenten: Die Wissensbasis und die Managementkomponente als Webapplikation, die als Server-side Java im Browser ohne weitere erforderliche Plug-ins läuft. Die Wissensbasis ist auch losgelöst von der Webapplikation frei im Web verfügbar unter <https://data.wiz.zpid.de/>. Die Webapplikation wird nach Fertigstellung Ende 2017 vom ZPID betrieben werden und steht der psychologischen Fachgemeinschaft für ihre Projekte nachhaltig zur Verfügung (einschließlich dem Speicherplatz für die Archivierung der Forschungsdaten). Darüber hinaus wird der Quellcode der Webapplikation nach Freigabe auf GitHub unter der Adresse <https://github.com/ZPID/DataWiz> bereitgestellt werden. Durch diese Bereitstellung können (1) Installationen von DataWiz in der eigenen Institution autark implementiert werden, (2) Anpassungen des Systems an die Bedürfnisse anderer Disziplinen erfolgen oder (3) einzelne Module, wie die Import- und Exportfunktionalitäten von SPSS-Dateien, unabhängig vom Gesamtkontext des Projekts nachgenutzt werden.

Zusammenfassend wird mit DataWiz die Kultur einer an Nachhaltigkeit, Offenheit und gemeinschaftlicher Nutzung orientierten Praxis des Umgangs mit Forschungsdaten gefördert, gestärkt und verankert. DataWiz ist somit ein Baustein der allseits geforderten Open-Science-Praktiken und trägt der Forderung nach Wissenschaftskommunikation als „Open Digital Scholarship“ Rechnung (Castro et al, 2015).

Hintergrund für das Verfassen dieses Beitrages ist das Poster „Integrating Open Science Practices into the Research Process in Psychology“, das vom Programmkomitee der *Open Science Conference 2017* (Berlin, 21.-22. März 2017) in einem *Open-Review*-Verfahren auf dem zweiten Platz gelandet ist.<sup>6</sup>

**Deskriptoren:** Psychologie, Datendokumentation, Forschungsdokumentation, Open Science

**Danksagung:** Der Autor dankt Ina Dehnhard, Armin Günther und Judith Tinnes für die kritische Durchsicht des Manuskripts.

<sup>6</sup> [http://www.open-science-conference.eu/call-for-project-presentations/\[13.3.2017\]](http://www.open-science-conference.eu/call-for-project-presentations/[13.3.2017]).

## Literatur

- Baker, M. (2015, August 27). Over half of psychology studies fail reproducibility test: Largest replication study to date casts doubt on many published positive results [News]. *Nature*. doi:10.1038/nature.2015.18248.
- Baker, M., & Dolgin, E. (2017, January 19). Reproducibility project yields muddy results: An ambitious effort to replicate cancer studies is provoking controversy [News]. *Nature*, *541*, 269–270. doi:10.1038/541269a.
- Castro, A., Hastings, J., Lange, C., Stevens, R., & Weichselgartner, E. (2015). Dagstuhl Manifesto: Digital scholarship and open science in psychology and the behavioral sciences. *Informatik Spektrum*, *38*(6), 572–573. doi:10.1007/s00287-015-0936-3.
- Chalmers, I., & Glasziou, P. (2009). Avoidable waste in the production and reporting of research evidence. *Lancet*, *374*(9683), 86–89. doi:10.1016/S0140-6736(09)60329-9.
- Dehnhard, I., Weichselgartner, E., & Krampen, G. (2013). Researcher's willingness to submit data for data sharing: A case study on a data archive for psychology. *Data Science Journal*, *12*, 172–180. doi:10.2481/dsj.12-037.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLOS ONE*, *4*(5), Artikel e5738. doi:10.1371/journal.pone.0005738.
- Macleod, M. R., Michie, S., Roberts, I., Dirnagl, U., Chalmers, I., Ioannidis, J. P. A. ... Glasziou, P. (2014). Biomedical research: Increasing value, reducing waste. *Lancet*, *383*(9912), 101–104. doi:10.1016/S0140-6736(13)62329-6.
- OECD. (2004). *Declaration on Access to Research Data from Public Funding*. <http://acts.oecd.org/Instruments/ShowInstrumentView.aspx?InstrumentID=157> [13.32017].
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLOS ONE*, *2*(3), Artikel e308. doi:10.1371/journal.pone.0000308.
- Pritsker, M. (2013, November 11). Opinion: Video saved the scientific publication [News]. *The Scientist*. <http://www.the-scientist.com/?articles.view/articleNo/38082/title/Opinion-Video-Saved-the-Scientific-Publication> [13.32017].
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, *86*(3), 638–641. doi:10.1037/0033-2909.86.3.638.
- Shekman, R. (2013, December 9). How journals like Nature, Cell and Science are damaging science. *The Guardian*. <https://www.theguardian.com/commentisfree/2013/dec/09/how-journals-nature-science-cell-damage-science> [13.32017]
- Vines, T. H., Albert, A. Y. K., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., ... Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*(1), 94–97. doi:10.1016/j.cub.2013.11.014.
- Weichselgartner, E. (2008). Fünf Jahre Primärdatenarchivierung in der Psychologie: Ein Erfahrungsbericht. In M. Ockenfeld (Hrsg.), *Verfügbarkeit von Information* (S. 259–267). Frankfurt a. M., Deutschland: DGI.
- Weichselgartner, E., Günther, A., & Dehnhard, I. (2011). Archivierung von Forschungsdaten. In S. Büttner, H.-C. Hobohm, & L. Müller (Hrsg.), *Handbuch Forschungsdatenmanagement* (S. 191–202). Bad Honnef, Deutschland: Bock und Herchen. urn:nbn:de:kobv:525-opus-2397.
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728.
- Wojcik, D., & Michaels, P. (2015). A taxonomy to support the statistical study of funding-induced biases in science [version 1; referees: 2 not approved]. *F1000Research*, *4*, Artikel 886. doi:10.12688/f1000research.7094.1.



**PD Dr. Erich Weichselgartner**  
Leibniz-Zentrum für Psychologische  
Information und Dokumentation (ZPID)  
Universitätsring 15  
54296 Trier  
[wga@zpid.de](mailto:wga@zpid.de)

PD Dr. Erich Weichselgartner ist stellvertretender wissenschaftlicher Leiter des Leibniz-Zentrums für Psychologische Information und Dokumentation (ZPID) in Trier. Er ist habilitierter Psychologe mit zusätzlichem Schwerpunkt Informatik. Seine Stationen vor dem ZPID waren u. a. die New York University und die Fraunhofer-Gesellschaft. Er interessiert sich für Informationsverarbeitung durch Mensch und Maschine und lehrt an den Universitäten Trier und Regensburg.