



Anreizsysteme: Datenpublikation und wissenschaftliche Reputation

Claudia Oellers, Denis Huschka

Workshop Forschungsdaten,
Arbeitskreis Forschungsdaten der Leibniz-
Gemeinschaft am 10. Mai 2012, Berlin

Gliederung

- Problem: (anders, als der Titel unseres Beitrages vermuten lässt) Datenpublikation bisher **ohne** Verankerung im wissenschaftlichen Reputationssystem
- Was sind die Implikationen der fehlenden Anreize?
- Möglichkeiten der Anerkennung von Forschungsdaten als wissenschaftliche Leistung: Welche Initiativen gibt es?
- Ausblick/ zukünftige Herausforderungen

vielfältige Argumente für Data Sharing:

- öffentlich finanzierte Daten
- Beantwortung neuer Fragestellungen
- Reproduzierbarkeit von Daten (Daten, die „irgendwo“ abgelegt sind, bedeuten nicht replizierbare Ergebnisse)
- Verbesserung wissenschaftlicher Ergebnisse (mehr Daten = mehr Erkenntnis)

Grundsätze zum Umgang mit Forschungsdaten

„**Qualitätsgesicherte Forschungsdaten** bilden einen **Grundpfeiler wissenschaftlicher Erkenntnis** und können unabhängig von ihrem ursprünglichem Erhebungszweck vielfach Grundlage weiterer Forschung sein. Dies gilt namentlich für die Aggregation von Daten aus unterschiedlichen Quellen zur gemeinsamen Nutzung. Die nachhaltige Sicherung und Bereitstellung von Forschungsdaten dient daher nicht nur der **Prüfung früherer Ergebnisse**, sondern in hohem Maße auch der **Erzielung künftiger Ergebnisse**. Sie bildet eine strategische Aufgabe, zu der Wissenschaft, Politik und andere Teile der Gesellschaft gemeinsam beitragen müssen.“

(Allianz der deutschen Wissenschaftsorganisationen, Grundsätze zum Umgang mit Forschungsdaten, 24. Juni 2010)

aber...



Nature 461, 160-163 (2009) |
doi:10.1038/461160a, published online 9
September 2009



Data Sharing is a bit like going to
the dentist. We can all agree that
it is a good thing to do...(Felicia
LeClere, 2010)

Gründe, Daten nicht weiterzugeben

- Forschungsdaten **nicht** verankert im wissenschaftlichen Reputationssystem
- Eine Publikation in einer Zeitschrift (möglichst mit hohem Impact Factor) ist eine wissenschaftlich anerkannte Leistung und findet Eingang in bibliometrische Indices. Auf deren Basis werden Entscheidungen über Karriereverläufe, Berufungen, Drittmittel gefällt. Dies gilt hingegen nicht für die Erstellung und Aufbereitung von Forschungsdaten.
- →Kein Anreiz aus forschungsökonomischer Sicht, in Aufbereitung und Verfügbarmachen zu investieren
- →Forschungsdaten sind im Vergleich zu Publikationen „**second class citizen of scientific records**“ (Adam Farquhar, 2009)

Gründe, Daten nicht weiterzugeben

Ergebnis
verschiedener
Befragungen
(European
Commission, 2012,
PARSE.Insight,
2009, Tenopir et al,
2011, Feijen 2010):

- Nur ca. 30 % der Befragten gaben an, offenen Zugang zu Ihren Daten zu gewähren (wenn möglich).
- Als Gründe wurden genannt: Fehlen von Zeit und Geld (Tenopir, 2011), rechtliche Unsicherheiten und Angst vor „misuse“ (PARSE.Insight).
- Aber fast alle gaben an, dass sie sich „credits“ für die Verwendung ihrer Daten wünschen!

„Credit where credit is due“ (Editorial, Nature, Dec. 2009)

- Datenproduktion als wissenschaftliche Leistung anerkennen (Empfehlungen des RatSWD, 2010, „Credit where credit is due“)
- Voraussetzung um wissenschaftliche Reputation zu erhalten, ist die Zitation von Daten.
- Notwendig sind Standards für die Zitation von Daten, vergleichbar mit der Zitation von Publikationen.

Zitation von Forschungsdaten

- Notwendigkeit der Etablierung von Standards, wie Datensätze zitiert werden sollten.
- Zitationen nicht als Anmerkung im Text, sondern in der Referenzliste („full citation“).
- Etablierung von Metadatenstandards
- Verwendung von Persistent Identifiers

Anerkennung von Forschungsdaten als wissenschaftliche Leistung: Was ist erforderlich?

"Der Wissenschaftsrat fordert insbesondere die Forschungsförderer auf, Anreize zu schaffen, um qualitativ hochwertige Daten zu archivieren und langfristig zu sichern. Zu diesem Zweck sollten Referenz- und entsprechende Zitationsmöglichkeiten für Datensätze aufgebaut werden. Persistent Identifier (PID) bzw. Digital Object Identifier (DOI) erlauben eine eindeutige Identifizierung und Zitierbarkeit digital hinterlegter Daten selbst dann, wenn sie ihre Speicherorte (in der Regel referenziert über den Uniform Resource Locator, URL) wechseln." *Empfehlung zu Forschungsinfrastrukturen in den Geistes und Sozialwissenschaften, Wissenschaftsrat, Berlin, 28.01.2011, S.58*

Anerkennung von Forschungsdaten: Akteure und Mechanismen

- Data Cite als internationales Konsortium mit dem Ziel, Forschungsdaten auffindbar und zitationsfähig zu machen
- Forschungsförderer: DFG „Sicherung guter wissenschaftlicher Praxis“ (1998) sowie seit 2010 berücksichtigt in der Antragstellung, NSF: Datenmanagementpläne
- Verlage: Data Journal Policies (submission guide lines)

Zusammenfassung Datenzitation

Datenzitation ermöglicht:

- Replikation wissenschaftlicher Ergebnisse
- Verbesserung von Forschungsstandards
- Credits für Datenproduzenten

Veränderungen der Zitations- und Publikationskultur

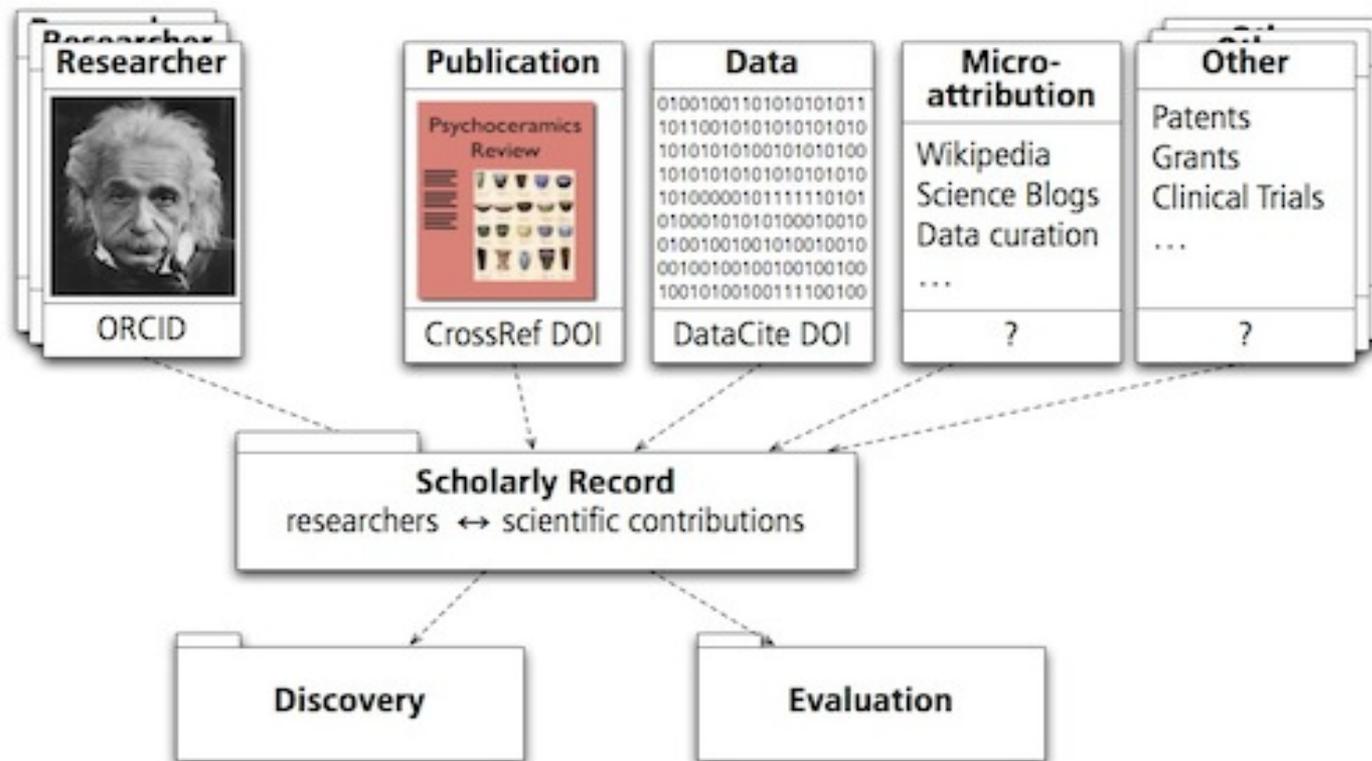
- Kritische Masse muss erst noch erreicht werden.



Veränderung der Zitations- und Publikationskultur

- PIWOWAR (2007): Verfügbarkeit des Datensatzes geht einher mit signifikant positiven Auswirkungen auf die Zitationsrate. Ähnliche Ergebnisse auch PANGAEA.
- Möglichkeiten der Veröffentlichung von Forschungsdaten → als Publikationen in peer-reviewed Journals - Beispiel European Data Watch oder Earth System Science Data (ESSD)

...aber auch Bewertung wissenschaftlicher
Leistungen muss sich ändern



<http://blogs.plos.org/mfenner/2011/08/28/on-microattribution>

Ausblick – offene Fragen

- Welche Leistung soll welchen wissenschaftlichen „credit“ erhalten?
- Wie lässt sich wissenschaftliche Leistung adäquat messen? Impactmessung?
- Wer erntet? Wer ist Autor?
- Frage der Versionierung/ Granularität
- Wie kann Datenverfügbarkeit im Rahmen kleinerer Projekte ohne Infrastruktur sichergestellt werden



- **Vielen Dank!**